

-899-

Identities = 147/531 (27%), Positives = 261/531 (48%), Gaps = 26/531 (4%)

Query: 19 IRICKLTKMSIRRVRYKQYLKQBZADKRIEFIEECSTKGLAGKLRLLAPKQWLE 78
 I + K K+I+ R ++Y+++ + IE+IE+ T G K+L QK WE
 5 Sbjct: 16 IEINKYMRKTIQKQIRIHKKYIYKRDVQAIIEWIEDNFYLTGNNMKLIKHPQKYNYR 75

Query: 79 TTWGFYHTVEVTKINPDTLSEYTDYBERLLIHEVPILVPRGTGKTLLSALAEVGGQIIDG 138
 G+ D ++E + LI+E+ + + RG+GK+L + I+ G
 10 Sbjct: 76 LMLGY-----DMVDEKG--VQVNLNKEIFLNLGRSGSGSLMATRVLAWMILGG 122

Query: 139 EWGADIQLLAYSRSQGYLFNASRAMLGNBSLLHYMREADIIRSTKQGLIYEITNSLMS 198
 ++G+ ++AY QA ++F+ R ++L Y E I +STKQG+ + +
 Sbjct: 123 QYGGESLVIAVDNTQARHVFQVRNQTEASDILRVY-NENKIFKSTKQGLETAFTKTFK 181

Query: 199 IKTSDYBSLDCNAHYNI FDEVHTYDDDFIKVVDGSSRRKKNMITWYISTWGTIKRDLF 258
 +T+D G N+ NIFDEVHTY +D + VN GS +K+ NW + YI+ G KRD L+
 15 Sbjct: 182 KQTNVTLRAQSGNSLNI FDEVHTYGEDITESVNGKSRQKQDMQSIYITSGGLKRDGLY 241

Query: 259 DKYNIWIDILDDKIINDSVMPMIYQLDVSEIHPDMQKAMPILGITTETKIARDIE 318
 DK + +++ ND +Y L++ ++ D W A+PL+G + + + E
 20 Sbjct: 242 DKLVERFKS--EEEFYNDRSPGLLYMLNHEQVKDKKNWIALFLIGDVPKMSGVIEYE 299

Query: 319 MSKNDPAQQALMAKTFNLFPVNNYLAIFYNSKECKGMSDKFDBSLFVGDDERNARCVID 378
 +++ DPA Q + +A LP+ + YF+ + K +P+ S+P R +GID
 25 Sbjct: 300 LKQGDPAQKFLAPNWLPMQDTAYYFTPDTK--LTFNLSVF-----NKNRTVIGID 352

Query: 379 LSDVNDICISPMVVRGEERHYLNKKFMPHITETLPEKELADKYTBWELSGMLHVHELDY 438
 LS + D+ ++SF+ + + F R E L E ++ +TE+ GL + + + Y
 30 Sbjct: 353 LSLIGDLTAVSPVCELEGKTYSHLTFTSVRSQYEQLDTEQQELNWFVDRGSLILLIDTEY 412

Query: 439 NDQAYIFEEELQPMDSNRILPFAVGVDYRINARELIRLFNDYDIGHIDIPQTVK---SLS 495
 + + + F S +GYD L L Y+ D D + + S++
 35 Sbjct: 413 INVADLIPYINDFRSKTGCLRLKIGYDPARYELLGLIERYFDPKGDINQRAIQGFSGM 472

Query: 496 NPLKVKYKAKMKGIIFDDPVATVNHANVRVKIDANNINIFENKAKKSID 546
 + +K+ K K K+I + V W N VKI + + K+ K+KID
 Sbjct: 473 DYIKLKLSKLVENKLIHNQKVMQALNATVAVKIGQSGDYMTYKLEKDKID 523

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 813

A DNA sequence (GBSx0862) was identified in *S.agalactiae* <SEQ ID 2469> which encodes the amino acid sequence <SEQ ID 2470>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3319 (Affirmative) < succ>
 50 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AB41469 GB:L35061 orfL4 [Bacteriophage phi-41]
 Identities = 86/374 (22%), Positives = 166/374 (43%), Gaps = 38/374 (10%)

Query: 12 FARIFRPNRRKSTRYLQRSISYNRRNSIYLDNYIKLSTDTAQLRPFKHUKITRNPQGV 71
 F+R N+ + + + Y S++ NI+NKI+ + + F HYK ++ G D
 55 Sbjct: 10 FSRGKLNDTQRVTAMQNEAVEY---TSAPVTINHNKIANETIKYFENHVICKYKSDVGS 66

Query: 72 SMWYYSDBLAEVLIVSPNPLEVPPVPSVIVTRMLADGVAVVPRW--KNGRLVEIWL 129

-900-

```

      +++   SDL EVL S       + FW V + +L       + P + K G LV+ + A
Sbjct: 67 TLISMAIGSLLDLVLNWSKGRKSRNMFQWQVKKILITRYIDLYPIFRKTGDLVDLLFA 126

5   Query: 130 KKTVTWTAESVELMLDDVAVELFLTDVWVFNPKLVTAQLNQITELIDINLALTEKLS 189
      + E + ++ +           + N+ T +D L + KL
Sbjct: 127 DNKKRYKPEELVRLISFPYI-----NEDTSILDNALAGIQITLE 165

      Query: 190 DQNSLRIGFLKLPFT---KADEHLKQOARDVDSMLDIANKGQIAYLBQGHSPQELSKDY 246
      G +G LK+ D+ K +A + +H +++ G+ + E EL KTY
10  Sbjct: 166 QGK--MKGLKINAFIDTNDQGFQDKAMLTIKNMQEMSNYNGLTPTDNKTEIVELKQDY 223

      Query: 247 STASKELEFLKSLQNLNAHGINEKLFCTDYETBOYRAYSSVMKLYQRYVSEHNKVKYPT 306
      S +K+E++ +KS+L + +NE + ++EQ +Y+S + +E+ K +
15  Sbjct: 224 SVLNKDEIDLKSELLTGYFMNENILLGTASQEQIYFFNSTIIFLLIQLEKELTYKLI 283

      Query: 307 KTKR--TQGN---KLAVFFDMALMISFKDLVBGGFKSYAGLNNSENFRETYLGLPFG 360
      R +GN +++V + + K+L++ ++ + N+ +G E
20  Sbjct: 284 TNRKRVVGNLYYERITVDNQLKFATLKEIDLYHENINQPIPTQQLL-VKMQEPIE 342

      Query: 361 GSEVFETINLNAVRI 374
      GG+V+ NLNAV +
20  Sbjct: 343 GSDVYIANLNAVAV 356

```

No corresponding DNA sequence was identified in *S.pyogenes*.

- 25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 814

A DNA sequence (GBSx0863) was identified in *S.galactiae* <SEQ ID 2471> which encodes the amino acid sequence <SEQ ID 2472>. This protein is predicted to be a prohead protease. Analysis of this protein

- 30 sequence reveals the following:

```

Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
35  bacterial cytoplasm --- Certainty=0.3496(Affixmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

40  >GP:AAF31089 GB:AP069529 protease [Bacteriophage HK97]
    Identities = 52/142 (36%), Positives = 73/142 (50%), Gaps = 11/142 (7%)

    Query: 21 FEAVASTYINTDREGDVMAKSCFDNLTLSKA-VVPMCLAHDR-NCVIGKHE-LSVDEKGL 77
    FE YAS ++NTD +GD++ G F N L ++ V M NH +GK +L+ DEKGL
45  Sbjct: 26 FEGYASVFNNVDSGDIIILPGAFKMLANQTRKVMANFFNHKTVKELVGVKWSLASEDEKGL 85

    Query: 78 RTGRTFNLSDEPAKTYTDLMMKSGALDSLSIGFFI--KDYRPIIDAKQPYGWIFKEVS-IF 134
    R A M+ G ++ +S+GF + DY I G IFK ++ +
50  Sbjct: 86 YVRQQLTCHSGAADLKAAMQHGTVKMSVGFSAKDDVYTIPT-----GRIFKNQIAIR 140

    Query: 135 EISVVTVPANPQATVDNIKEFD 156
    EISV T PAN QA + +K D
    Sbjct: 141 EISVCTFPANQGLAAMKSV 162

```

- 55 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 815

A DNA sequence (GBSx0864) was identified in *S.agalactiae* <SEQ ID 2473> which encodes the amino acid sequence <SEQ ID 2474>. Analysis of this protein sequence reveals the following:

```

Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2247 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10155> which encodes amino acid sequence <SEQ ID 10156> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC27185 GB:AF009630 16 [bacteriophage b11170]
Identities = 70/249 (28%), Positives = 121/249 (48%), Gaps = 23/249 (9%)

Query: 51 LEQLKTDASLSVSQATA--IKETIAGLDSIDIBETEELSK-AAKIHK-----EKOK 98
      L +LK + SL SQ +K I L ++R E+ Ls+ + +IHK EK K
Sbjct: 13 LAELKENNVSLKSQINGFEVKNALIEDLPK-VQELERTLSNSIETIKIENELNAQEEKPK 71

Query: 99 GNTFM-DYLKTKAALDFVRILMUNBGSANSARKAWENLVKGV--TNLTKLPEPFLI 155
      G M +++++ A +F +L N G + +AW A L E GV T+ T LP ++
Sbjct: 72 GKAKGNTNFIESQNAVTEFFDVLKINSKGE-IRNMMALAEANGVTITDTTFLPKRLVE 130

Query: 156 AIQDAFTNYNGILN--HVSXOPRYAVRVALQTQVSQAKQHKAGTKKQEDFTFLDPTINS 213
      +I A N N + HV+ V + + ++A+ HK G+TK ++ T T+
Sbjct: 131 SINTALLNTINPVFVPHVTNVGALLVSRFSOSS-AEAQVHKDQTKTEQAATLTIDTLEP 189

Query: 214 ATVY-IKYAFEYSLDKKDTTGAYFNYVMKELAQCFI-RTIERAVVIGDGKSN-SAEKKT 270
      VY ++ E + + +N ++ EL Q + + + A+V GDG + + DK
Sbjct: 190 VMVYKLSLAERVERLQMSYSELNLTVAELTQATVNVKIVDLALVEGDGSGNGFKSIDKEA 249

Query: 271 EIKSIABET 279
      ++K I + T
Sbjct: 250 DVKKIKKIT 258

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 816

A DNA sequence (GBSx0865) was identified in *S.agalactiae* <SEQ ID 2475> which encodes the amino acid sequence <SEQ ID 2476>. Analysis of this protein sequence reveals the following:

```

Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3068 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 817

A DNA sequence (GBSx0866) was identified in *S.agalactiae* <SEQ ID 2477> which encodes the amino acid sequence <SEQ ID 2478>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0437(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 818

A DNA sequence (GBSx0867) was identified in *S.agalactiae* <SEQ ID 2479> which encodes the amino acid sequence <SEQ ID 2480>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3181(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10153> which encodes amino acid sequence <SEQ ID 10154> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 819

A DNA sequence (GBSx0869) was identified in *S.agalactiae* <SEQ ID 2481> which encodes the amino acid sequence <SEQ ID 2482>. This protein is predicted to be a major structural protein. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3364(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```


The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA74331 GB:133769 unidentified ORF28; putative [Bacteriophage
      b1L67]
Identities = 55/201 (27%), Positives = 84/201 (41%), Gaps = 18/201 (8%)

5  Query: 9  EVTHGNANGF-YAKIAKTDAGALDLQKYPPTGLRSTSYETSQBSNAYAD-NVEHVLQ 66
      E+THG  G + + + G  P GLR  ++ QR+  +YA N + +
      Sbjct: 8  ELTHSLGTGVVPTDLTGSKTGI-----PLASLRGIETDSKQENKNFYAGFNAPRTIA 60

10  Query: 67  GKSTEGSITTYQIPKQFMIDHLGKKLINSTFPALIDTGWIN-PWGYARTVIDEFGAR 125
      G K T+  + +Y +P F  LG  S  L D N  + + YAR  D+ G
      Sbjct: 61  GAKLTQ1KVKSTDLPODFATHLG---FGSVQGLTDDVANYKPYGFATERYDELDGTG 117

15  Query: 126  IEEFIHWITNVKASAPKSGSTSTDETSATPKKIRIPCTASPNFVDSHKKPVSEIWRDGS 185
      + + + +V+A+ P +  DE S T KE E  T + + F + + K+  + D
      Sbjct: 118  YKA-TFFPSVQATTPSDIAEADERSPTGKEYEHEATVTTGDFTLGDKRLPVKFKVSDTE 176

      Query: 186  KGI-VRGK---FKLFDADKSP 202
      T  GK  F KLP D P
20  Sbjct: 177  LATGTSKHALAKKLPIDLPK 197
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 820

A DNA sequence (GBSx0870) was identified in *S.agalactiae* <SEQ ID 2483> which encodes the amino acid sequence <SEQ ID 2484>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

30  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2531(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 821

A DNA sequence (GBSx0871) was identified in *S.agalactiae* <SEQ ID 2485> which encodes the amino acid sequence <SEQ ID 2486>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

45  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2972(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 822

A DNA sequence (GBSx0872) was identified in *S. agalactiae* <SEQ ID 2487> which encodes the amino acid sequence <SEQ ID 2488>. Analysis of this protein sequence reveals the following:

Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3860 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 823

A DNA sequence (GBSx0873) was identified in *S. agalactiae* <SEQ ID 2489> which encodes the amino acid sequence <SEQ ID 2490>. Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -14.22 Transmembrane 605 - 621 (569 - 631)
INTEGRAL Likelihood = -8.12 Transmembrane 583 - 599 (569 - 604)

----- Final Results -----

bacterial membrane --- Certainty=0.6689 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB70053 GB:AF011378 unknown [Bacteriophage sk1]
Identities = 159/709 (22%), Positives = 285/709 (39%), Gaps = 112/709 (15%)

Query: 128 SILNLAKLELDNVAKELDIVNQLKLELDPNVSLARQMKLAGKQSELAGDKVQVEKKQQA 187
S+ +N + + E + L:LDP N + Q K L Q L+ DK +Lk+ ++
Sbjct: 21 SLKGVNTAMSGLRGEAKNLDALDKLDPTNTQKAGLQKILQTOGLSRDKATKIKQELSS 80

Query: 188 LGDEK-IGTEEMRQLNQIQAEVHVLEKIDRAMDILGESSRATGDI--KEATSYLRADV 244
+ G +W QL ++G AE + +++ + + S + DI K T + ++
Sbjct: 61 VDKSSPAGQKKWLQIRDLGTAFQANRLEGEIKQVGAISGSGWIDAKMDITGVNSGI 140

Query: 245 MMDVADKAG-----GIQKRMVDAGKMTVDWASEIDEALDVTTKGLTGD----- 289
+ +G OIG V A + W + +A+DT L
Sbjct: 141 DGMKSRFSGLRSTAVGVPRIQSSAVSAGNGLKGM--VSDAMDTQAMISLQNTKFKG 198

Query: 290 -----ALAEHQEIADTIATG-----MPTSPQAGD-----AVGEL-----NTQFGLT 326
+ Q +AKD + T+F GD ANG+ N H3 T
Sbjct: 199 NGQDFDYVSKSMQTLAKDTNANTEDTLKLSITFIGLGDSAKTAVGTALVKANQAGGT 258

Query: 327 GEAKGASRLL-----IKVPEINE-TD-----ISSATSARQAEAYG--LIRAK 367
GE+LK + + IN+ TD + S+ + A++ YG +A
Sbjct: 259 GEQLKGVVQAYQMGAGCKVSAENINQLTDNNTALGALKSTVMBNPAKLYQSGPASAS 318

Query: 368 DLGMV----LDNVTKAAQDTGQSVDTIVQKALDGAPOIKIGLSLPEBGA-----ALIGK 417

[illegible]

-907-

930 940 950

2886 2916 2946 2976 3006 3036 3066 3096

LGAIGQSIANTYNTSSNNINVNPSGVITREKADLMRLANVGNKKIASELORKTNILRGMA*OKSMNLFIV*KHLLSVNY
| : | :: : | : | | | : | : | : || :
LGSGGYGLSTNSVSSDNRITYTTFNVGGAGQDVSNLARIRREFELORA

960 970 980 990

SEQ ID 8664 (GBS58) was expressed in and purified from *E.coli* as a GST fusion. The purified protein is shown in lane 10 of Figure 193.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 824

A DNA sequence (GBSx0874) was identified in *S.galactiae* <SEQ ID 2493> which encodes the amino acid sequence <SEQ ID 2494>. Analysis of this protein sequence reveals the following:

Possible site: 43
>>> Seems to have no N-terminal signal sequence

```

----- Final Results -----
20      bacterial cytoplasm --- Certainty=0.2732(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

25 No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 825

A DNA sequence (GBSx0875) was identified in *S. agalactiae* <SEQ ID 2495> which encodes the amino acid sequence <SEQ ID 2496>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence
```

```

35      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2467(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10151> which encodes amino acid sequence <SEQ ID 10152> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10935> which encodes amino acid sequence <SEQ ID 10936> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2497> which encodes the amino acid sequence <SEQ ID 2498>. Analysis of this protein sequence reveals the following:

```

45      possible site: 40
      >>> Seems to have no N-terminal signal sequence

```

----- Final Results -----

-908-

bacterial cytoplasm --- Certainty=0.2136 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 55/240 (22%), Positives = 92/240 (37%), Gaps = 20/240 (8%)

```

Query: 4 INELTIDGVKTSSFKCDVLVETRPNVIVSS--KTALLEHDGISGAVVQSNRHHGLIEKP 61
      I ++ ID TSS VL I+S S + +G S + N + I
Sbjct: 2 IPKVILDDFTSSIPKCVLTIGYDVLGILSPSFVENRAYGMNKTSSRRLSSTNIESKPTIM-- 59

Query: 62 YHITLIEPSDESIYRFSALLNREKFW--LENEQEPTIRLMWCYKVSFEIGKDEFGAMVVVD 120
      +H++ ++ I L + +FW + N ++ Y S +I +N V +
Sbjct: 60 WHLSTFDDAVNLINHLGLSKKIEFWHPNS-----IYYDCLSVKINAVTMSSNRVTL 113

Query: 121 TFICHPTKFKFTTIDITLITGNGVLRVQGSALAPKAITVVGQASSETSPTIGKQVLEKEL 180
      +P ++ K + GNG + G+ + PKI V G + + TIG QV+L L
Sbjct: 114 KLALYPPRYAKGVSDVVIAGNGVNNAGNVSEPKIVVHG--TGRGTITLIGKQVLEL--NL 170

Query: 181 SESLVMINDEPNPSFKTASGKL---IKWAGDFITVDTARGQNVGVVLGAGITSLKPFIVW 237
      S + + A G + I+ G F + +G+ + GIT W
Sbjct: 171 SGKATIECKHGQQCVYDABGNVNSIRIGSFPEIQPG---TQGIAGVSGITRTITISPKW 227

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 826

A DNA sequence (GBSx0876) was identified in *S. agalactiae* <SEQ ID 2499> which encodes the amino acid sequence <SEQ ID 2500>. This protein is predicted to be PblB. Analysis of this protein sequence reveals the following:

```

30 Possible site: 27
    >>> Seems to have no N-terminal signal sequence
        INTEGRAL Likelihood = -0.00 Transmembrane 952 - 968 ( 952 - 968)

35 ----- Final Results -----
        bacterial membrane --- Certainty=0.1001 (Affirmative) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

40 >GP:AMG16640 GB:AY007505 PblB [Streptococcus mitis]
    Identities = 145/542 (26%), Positives = 255/542 (46%), Gaps = 52/542 (9%)

45 Query: 1 MLPLLRNVRTVKNQNPILHEASSAIVKEETNGDFYLTVRYPTTDSGIYQLKEDMLIKS 60
      M++L + N PL+ A + + +E N + LT R+P +D +++ +KE+ +K+
Sbjct: 1 NIYLTNGNT-----PINAAYADKISQGANSTYQLTFRPPTSUV-LWEKLEETFLKA 51

Query: 61 PVFVLQQLPRIKKPIENDSDMTIYHVSDDIMKRSITFPVSVGQGCAMALSQMVAQNAK 120
      + G Q F I + + + + A V + I P+S+ + ALG+ +
Sbjct: 52 D-DLHGEGQDFVIFSVQKKGNYIQVYANQVMILNINYPINPISLDRATGSTATSRFPAGSI- 109

Query: 121 TGLGDFSPFSDIMDSRTFTFTTETILYSVLMDCKHSIVGTWEGELVDNFPALSIKRSRGA 180
      T FSF SDI + TET + + D KHSI+G W G+LVR + + + + G+
Sbjct: 110 TRYNTFSFPFSDIDSRHTFTNTDSVNAVAFIKD-KHSILAQKQGLDVRHGYQVRLLNKGS 168

55 Query: 181 DRGVVITHNKLGYQRTKNSQGVVTRIHARSTFKPDGAB--DEVTLVSVSDPLINSYFY 239
      + + KNL SYQ +++ + TRI ++T K +G + + V VDSPL+L N Y
Sbjct: 169 ENESLFWYKKNLSSYQHTSTFKSLKTRITFKATVKGEGKAPDRKFSVVVDSEPLVKNYSQ 228

60 Query: 240 INEKEYENNAFTVED--LRKWRARAKFTNEGIDKVSDAIEIYAYELDGQVWNLGDTVHLK 297

```

-909-

I E E N + + + L R K + E F D + D + E I + V + D V + L
 Sbjct: 229 IYEDVIEVDQVKEVDQLRKYGEVQVPTTLTCLMLRDSLEIQVSGKSDVVPVQIFDIVSLF 288
 Query: 298 SRKHSADLYKKAIAYEPNALITREYISITPDOKPGVGGSGVSSGLSN-VADAILVASATAQ 356
 + D + KK Y + + + + + SI F G SG + S + LSN V + DA + + Q
 Sbjct: 289 HDRFMDVRRKKITKYTVSPMAKKLSIGF---GQPKSGLSNMNLNAVSDAVNKVTHQLQ 344
 Query: 357 D---VAVQRAVGNANAAFDAPGKTKTKINDIRIAKAKVESFKSELSNRNMNQLLP--- 410
 + + + KNA + A D + + + D + AKAK E K L + + D +
 Sbjct: 345 GQFATQLGKEIKNADLAFDRKKEELVNOFTDGLNAAKAAAEVKKSLTETIDQRFDRFDS 404
 Query: 411 -----LATEAKMLASQAQADLTRKETELRAELNBOVSTENAVK 448
 LA EAK + + QA + + K E + + + TS +
 Sbjct: 405 TGLNEIKQKABEALQVGNATLLAQRAKQISEQAPQCMDSKFAEYKQSVDRPTSLSSQL 464
 Query: 449 ISLTNLSHNMMDLIKQKALNDLDAETRLKEADSVQQLATKRVEDKLTGLSTKLESFVSG 508
 NL + D + + + + L + E + D + + A + + L + S + VGS
 Sbjct: 465 AGKANL---IDFQVQKSNLYERIIIGSSSDIAEKVARMTLNLQFQVEVGKYS-AVGS 520
 Query: 509 YN 510
 N
 Sbjct: 521 PN 522
 Identities = 47/183 (25%), Positives = 83/183 (44%), Gaps = 22/183 (12%)
 Query: 867 VTTLRVTKGTIPADNSPSPDILKAYSDTKLEQTANEIKASVTSLOHKTLEKTVITMSRG 926
 + T L + GT N P + D + D I E T Q T + +
 Sbjct: 667 MTELDIFYEOTDRRNQPAPEADTLETDKILEAT-----QTKLTLQGS 709
 Query: 927 IVLRAGKTSNDVARAIGSYFKVTPEALFSLKGVGNMLVDGVSFSRKLVAFTGVTGH 986
 + TS A + I S T + I + + I + G L + D + + + G
 Sbjct: 710 FAIQ-NLTS---AGSIVSQINATNQLIEAKIRLRGKTLTD-ELTADIGYFKFLVGE 764
 Query: 987 VKGALITGVLLAAEAVTAEKLVQAPFNKLMANDALYKLFKSAFITQVQSVTISASQ 1046
 + + + + + TA + L + DQA + + + D + L AK AFI + + SV + SA +
 Sbjct: 765 GTFAKLNAEIIGSKTITADKLIMDQANRLFVSSDIFTDTLAARAFINKLRVSVVBSATL 824
 Query: 1047 ISG 1049
 G
 Sbjct: 825 FBG 827

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2501> which encodes the amino acid sequence <SEQ ID 2502>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2445 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 143/552 (25%), Positives = 251/552 (44%), Gaps = 43/552 (7%)
 Query: 11 TVKNGIPLHEASSAIVKRETNIDPFLTVRYPTDSCIVQLIKEDMLIKSPYVPLGAQLP 70
 + K + PL A + + E N D + L + YP LIK + + + + + + G + QLF
 Sbjct: 3 SIKDDNIPLVAAFPEDKITQEANSYKLNFKYPAKHE-YRPLIKKIIILEAD-DIHESQLP 60
 Query: 71 RIKKPIENDSDNDITAYHVSDDIMKRSITPVSFVQGCAMALSQMVONAKTGLDPSFIS 130
 RI + + + + A V + DD + I + SV + S + + + K PSF S
 Sbjct: 61 RIFRITKRGGYINVYANQVADLLNGVALDITISVDRVQGMVMSRELASGIRRE-IPPSFIS 119
 Query: 131 DIMDSRTNTTETETLYSVLMGKHSIVGTWEGELVRDNFALSIKRSRGADRGVITTHK 190
 DI TFN + + + L + GHSI + G W GELVR + + + + G D + K
 Sbjct: 120 DIDGRHTFNQSDVSVM-DALANGKHSIMGQWGKELVRNKYQINLLKKAGKIDTETLPMYKK 178

-910-

Query: 191 NLKSYQRTKNSQGVVTRIH-----ARSTFKPDG-----ARDEVILRVSVDSPLI 234
 NLKSY+ T +G+V+ +H + DG + + T+RVSV+ S L
 Sbjct: 179 NLKSYETDTIKGLVSLILHVAEVEEHEVETREASDQVCHSSES PKKRTI RVSVESK LK 238

5 Query: 235 NSYPYINKE--EYENNNAEIVEDLRMAEAKPTNEGIDKVSDAIEIAYELDQGVVNLGD 292
 +++P I EK + + + + + T EDL + + F D + + + + I + V L D
 Sbjct: 239 DTHPIIVKTIKVVQDQVKTRELLAYGKKYKEKILCDIPKSLKIDVTVNNYGAVALPFD 298

10 Query: 293 TVNLKSRKHSADLYKKALAYEFNALTEEYISITFTDKPGVGSQVSSGLGNVADAILVAS 352
 T + + DL + Y F + S I F G + + + + SN D + S
 Sbjct: 299 TAIVFHELYDRIDRMQITGYRFAPMANRLKSIIF----GEIKTNLAKQINQIDNKVABS 354

15 Query: 353 ATAQDVA----VQRAVKQNNAPFADRFKTKTKINDDIHIKAKVSEPKSELSNR-MDNQ 407
 D A +Q+ + NAN ED + K + + I D I+ A+A E +E+ + + +
 Sbjct: 355 TAQHDAAFEAKLQKQIDNANRIFDTKBAKLREIEDGIKKAANAENVKVAENAKVLENE 414

20 Query: 408 LLPLATEAK-----NLASQAQADLTKKIELRAHAKQVTSFAVKISLTLNLSHNDIYK 462
 L A + + + A + + D + K E R L + + + L + D +
 Sbjct: 415 ELAKAVDERLKKFISADATYEQDFDKLEEFRTSLKDLVDEKQIDDALAKGFSKDSLA 474

25 Query: 463 QKALNDLRDAETRLKEADSVQQL-ATKRVEDKLTGLSTKLESPVGGVYNNVIDGGSPKEL 521
 +ET A+ V T ++L G + K + P GY + GE E
 Sbjct: 475 DIKAKLEDTSSETATVTANIVSGTGGTYNNRRLOGDTRKVTPE-QGYIDIAHNGSGFE- 532

30 Query: 522 MANFYKTYIDIN 533
 GKTY I+
 Sbjct: 533 ----EGKTYTIS 540

A related GBS gene <SEQ ID 8665> and protein <SEQ ID 8666> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 3
 SRCFLG: 0
 MoG: Length of UR: 11
 Peak Value of UR: 1.54
 Net Charge of CR: 1
 MoG: Discrim Score: -3.43
 GVH: Signal Score (-7.5): -5.44
 Possible site: 58
 >>> Seems to have no N-terminal signal sequence
 Amino Acid Composition: calculated from 1
 ALON program count: 1 value: -0.00 threshold: 0.0
 INTEGRAL Likelihood = -0.00 Transmembrane 897 - 913 (897 - 913)
 PERIPHERAL Likelihood = 1.48 932
 modified ALON score: 0.50
 icml HYFID: 7 CFP: 0.100

45 *** Reasoning Step: 3

----- Final Results -----

50 bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

55 32.8/53.9% over 503aa
 E0AD[33685] hypothetical protein Insert characterized
 HQAD[71773]76294 hypothetical protein { } Insert characterized
 SP[P1537]YHYA_BPH44 HYPOTHETICAL 65 KDA PROTEIN IN HYALURONIDASE REGION. Insert
 characterized
 60 GP[215054]gb|AAA98102.1|X19348 ORF {Streptococcus pyogenes phage H4489A} Insert
 characterized
 PIR[B30566]B30566 hypothetical protein - phage H4489A Insert characterized
 ORF00870[1957 - 3777 of 4272]

-912-

Query: 370 AINLNSRGVQIAGKNIALDGMTT---VNGAF-----GAKLGEFI-----KLRAD 410
 AI L S +++G N+ +DG+ T V GA GA G + K+ D
 Sbjct: 897 AIALPSSLIKVSG-NMLWDGSSVTSRKLVTGAVEIGHVKAGAITGVLLAAEAATBKLKVD 955

Query: 411 QIIGGITDANKINVINLQASIVGLDANFIKARISYAIT-DLEGGKVIKARNGAMTIDLQ 469
 Q + AN + L A S FI S I+ + G VIK N AM I +
 Sbjct: 956 QAFFNKLMANDAYLKQLFAKSA-----FITQVQSVTISASQISGSGVIKALNANMETQW 1009

Query: 470 SQQINHYTNESAMRRIDSSTASQFIKMTKSGFISEIGMQAMIVIGSNSDGSNNHNT 529
 SQI +YT+++A++R+ S +QF+K +G +S GN A +TVIGSN G+3+ +
 Sbjct: 1010 SQILYLYTDAQALKRVLSQYPTQFVKFA-TGTVSGKGN--AGVTVIGSNRGVGTSTNDG 1066

Query: 530 FGGIRIYNGKSSYSTSFLVLAGN--RVAIYGNKNSPWLFDSTISGYAYLIPONDRIK 587
 F G+R WNG + ++LUG+ R+A N W + SG + P N
 Sbjct: 1067 FVGVRWNG---SNIDSULVGEIRLASSAFNSDGNVVRTLDSGLK-ITPIN----- 1116

Query: 588 HVIGPAERKIDQIHVGDIVV-CGERVAMMLKDL 619
 RA + +I VGD+++G L+D+
 Sbjct: 1117 ----RAAERNRRIEVDVWILKNGSYSSLRDI 1145
 Score = 31.3 bits (69), Expect = 0.038
 Identities = 34/151 (22%), Positives = 62/151 (40%), Gaps = 13/151 (8%)

Query: 160 QNADKKLSASYQLGIDGLKATMRSDKIGLOAEIOTTAQGLYORYNIRKLSAKITTTSS 219
 Q A K +A++ K + D +A++++ L R DN++ L+ + +S
 Sbjct: 306 QRAVQNVAVAFDAEPGKTKTKINDIEIANKVFSFKSELNRMDQQLPLATEANLAS 365

Query: 220 GTTAYESKLDGLKAEFT---SNQGMKVELES-----KISGLSTQQTARQISGE 268
 K LRAE S + ++ L + K L + A R + +
 Sbjct: 366 QAQADLTRKIELRAELNRQVSTBAVKISLNLNLSHMDIHKQKALNDLRDAETR-LKEA 424

Query: 269 ISNRGAVSRVQGLDSYQRLQS-ARGNYN 298
 S ++ A RV+ L +L+S + G YN
 Sbjct: 425 DSVQQLATKRVEDKLTGLSTKLBSFVGGYN 455

SEQ ID 8666 (GBS202) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 50 (lane 5; MW 132kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 827

A DNA sequence (GBSx0877) was identified in *S. galactiae* <SEQ ID 2503> which encodes the amino acid sequence <SEQ ID 2504>. This protein is predicted to be nuclear/mitotic apparatus protein. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2847 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 828

A DNA sequence (GBSx0879) was identified in *S.agalactiae* <SEQ ID 2505> which encodes the amino acid sequence <SEQ ID 2506>. Analysis of this protein sequence reveals the following:

```

Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3420 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 829

A DNA sequence (GBSx0880) was identified in *S.agalactiae* <SEQ ID 2507> which encodes the amino acid sequence <SEQ ID 2508>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have an uncleavable N-term' signal seq
INTEGRAL Likelihood = -7.54 Transmembrane 10 - 26 ( 2 - 28)

----- Final Results -----
bacterial membrane --- Certainty=0.4015 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP: CAB07984 GB: Z93946 hypothetical protein [bacteriophage Dp-1]
Identities = 67/136 (49%), Positives = 91/136 (66%)

Query: 1 MPPMLIDSTVVVMMVTVLGGLESTIIITTSANRKDQLIKHQYEDIKRDLGLIDKRVITIH 60
MP WL D+ V+ ++T G+ ++ K K EDI LS L +V ID
Sbjct: 1 MPMWLANDAVLTTIITACSGVLTVLNLKLPWKSNKAKSVLEDISITLTKQQVDGIDQ 60

Query: 61 TTTTETKISRITKDGTLKIQRYLPHDLTKELISQGYTTIEHPRELSILFESYQLLGNGE 120
TT ++ +DGT KIQRYL+HDL +E+ GYTT+HPRELSILFESY+ LGNGE
Sbjct: 61 TTVAINHQNDVIQDSTRKIQRYLYHDLKREVITGVTTLDHPRELSILFESYKILGNGE 120

Query: 121 IEALPEKPKQLPIRED 136
+EAL+EK+K+LPI E+
Sbjct: 121 VREALYKPKLPPIRED 136

```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 2508 (GBS118) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 32 (lane 5; MW 42kDa).

GBS118-GST was purified as shown in Figure 198, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 830

A DNA sequence (GBSx0882) was identified in *S.agalactiae* <SEQ ID 2509> which encodes the amino acid sequence <SEQ ID 2510>. Analysis of this protein sequence reveals the following:

```

Possible site: 53
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8667> and protein <SEQ ID 8668> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1   Crend: 5
McG: Discrim Score:      6.58
GVH: Signal Score (-7.5): -0.49
      Possible site: 53
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value: 12.15 threshold: 0.0
PERIPHERAL Likelihood = 12.15      84
modified ALOM score: -2.93

```

```

*** Reasoning Step: 3

----- Final Results -----
      bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear)

```

SEQ ID 2510 (GBS56) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 8; MW 9.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 21 (lane 10; MW 34.9kDa).

GBS56-GST was purified as shown in Figure 195, lane 7.

Example 831

A DNA sequence (GBSx0883) was identified in *S.agalactiae* <SEQ ID 2511> which encodes the amino acid sequence <SEQ ID 2512>. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 832

A DNA sequence (GBSx0884) was identified in *S. agalactiae* <SEQ ID 2513> which encodes the amino acid sequence <SEQ ID 2514>. This protein is predicted to be N-acetylmuramoyl-L-alanine amidase. Analysis of this protein sequence reveals the following:

```

Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0342(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB07986 GB:Z93946 N-acetylmuramoyl-L-alanine amidase
    [bacteriophage Dp-1]
    Identities = 96/141 (68%), Positives = 118/141 (83%)

Query: 1 MEINTETALAWMSARQKQVSYSDYRDPNSYDCSSSVYYALRSAGASSAGWAVNTEYMH 60
M ++ E +AWM AR+G+VSYSMD+RDGP+SYDCSSS+YYALRSAGASSAGWAVNTEYMH
Sbjct: 1 MGVDIEKGVAMWQARKGRVSYSDYRDPNSYDCSSSVYYALRSAGASSAGWAVNTEYMH 60

Query: 61 DWLIKNGYELIAENVDDNNAVRGDIATWGRHSSGAGGHVMPIDPENTIHCHNANNGIT 120
WLI+NGYELI+EN W+A RGD I W+G S+GAGGH MFID +NIHCN+A +GI+
Sbjct: 61 AWLIENGVELISENAPWDAKRGDIFWGRKGASAGAGGHTGMFIDSDNIHCNYAYDGIS 120

Query: 121 VNNTYNTAASGMYCYVYRL 141
VN+++ +G Y YVYRL
Sbjct: 121 VNHDERMYTAGQPYVYRL 141

```

No corresponding DNA sequence was identified in *S. pyogenes*.

A related GBS gene <SEQ ID 8669> and protein <SEQ ID 8670> were also identified. Analysis of this protein sequence reveals the following:

```

RGD motif 81-83

The protein has homology with the following sequences in the databases:

58.2/72.9% over 182aa

```

GP[1934766] N-acetylmuramoyl-L-alanine amidase (bacteriophage Dp-1) Insert characterized

```

ORF00875(301 - 1044 of 2004)
GP[1934766][emb]CAB07986.1|[Z93946(1 - 183 of 296) N-acetylmuramoyl-L-alanine amidase
(bacteriophage Dp-1)]
%Match = 15.5
%Identity = 58.2 %Similarity = 72.8
Matches = 107 Mismatches = 49 Conservative Sub.s = 27

```

```

      234      264      294      324      354      384      414      444
LQKYNTHMSDDLLTFVESAVKQMHDAWKE*PMEINTETALAWMSARQKQVSYSDYRDPNSYDCSSSVYYALRSAGAS
      :: : ||| ||:|:|||||:| | :|||||:|||||
      MGVDIEKGVAMWQARKGRVSYSDYRDPNSYDCSSSVYYALRSAGAS
                        10      20      30      40

      474      504      534      564      594      624      654      684
SAGWAVNTEYMHDLIKNGYELIAENVDDNNAVRGDIATWGRHSSGAGGHVMPIDPENTIHCHNANNGITVNNTYNTA
|||||:|||||:|||||:| | :||| ||| :|:||||| ||| :|||||:| :|:|:|:|:
SAGWAVNTEYMHDLIKNGYELIAENVDDNNAVRGDIATWGRKGASAGAGGHTGMFIDSDNIHCNYAYDGISVNHDERMY

```

-916-

```

        60          70          80          90          100          110          120
714      744      774      804      834      864      894      924
5  AASGWMYCYVYRLKSGASTQKSLDTLVKETLAGNYGNHSAKAVLGNQYEAVMSEVINGKTTNQKIVDQLAQVEVLGKH
   : | | | | | :
   YYAQGPYYVYRLTNA-----
               140

954      984      1014      1044      1074      1104      1134      1164
10 GNGRARKKSLGSGYDAVQKRVTELLKKQPSRPKQAQVNVKPIETKTSQTELTGQATATKRGDLSFGVITLKKAVLDKIL
   | : | | | | | : | : | | : : : : | : | :
   -NAQFAEKKLGWQKDATGFWYARANGTYPKDRFEYIEENKSWFYPFDQGYMLAEKWLKHTDGNWYFDRDGYMATSWKRI
               150          170          180          190          200          210          220

```

SEQ ID 8670 (GBS302) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 50 (lane 6; MW 55kDa).

The GBS302-His fusion product was purified (Figure 205, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 302), which confirmed that the protein is immunoreactive on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 833

A DNA sequence (GBSx0885) was identified in *S.agalactiae* <SEQ ID 2515> which encodes the amino acid sequence <SEQ ID 2516>. Analysis of this protein sequence reveals the following:

```

Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1509(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 834

A DNA sequence (GBSx0886) was identified in *S.agalactiae* <SEQ ID 2517> which encodes the amino acid sequence <SEQ ID 2518>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1264(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CA13473 GB:299112 similar to hypothetical proteins (Bacillus subtilis)
Identities = 25/68 (36%), Positives = 41/68 (59%)

```

-917-

Query: 4 IENLIAIVKPLISQPDQITIKIQGQPEFLEYHLDLDTQDIGRVIGKKGRTTITAIRSIVY 63
 +E+LI+ IV PL+ PD + + + + + L + D G+VIGK+GRT AIR+ V+
 Sbjct: 6 LEDLIHVHIVPLVDHPDIDIRVIREETDQKIALRLSVHKSDTKVIGKQGRATAKARTAVF 65

5 Query: 64 SVPTQGGK 71
 + Q K
 Sbjct: 66 AGVQSSK 73

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2519> which encodes the amino acid sequence <SEQ ID 2520>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.1012 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 72/79 (91%), Positives = 75/79 (94%)

Query: 1 MDTIENLIAIVKPLISQPDQITIKIQGQPEFLEYHLDLDTQDIGRVIGKKGRTTITAIRS 60
 MDTIENLIAIVKPLISQPD LTIKI+D F+PLEYHLDL QDIGRVIGKKGRTTITAIRS
 Sbjct: 1 MDTIENLIAIVKPLISQPDMLTIKIEDTFDFLEYHLDLQDIGRVIGKKGRTTITAIRS 60

25 Query: 61 IVYSVPTQGGKRVLIIDEK 79
 IVYSVPT GKKVRL+IDEK
 Sbjct: 61 IVYSVPTLGKKVRLVIDEK 79

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 835

A DNA sequence (GBSx0887) was identified in *S.agalactiae* <SEQ ID 2521> which encodes the amino acid sequence <SEQ ID 2522>. This protein is predicted to be ribosomal protein S16 (rpsP). Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.3654 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:BA006202 GB:AP001515 ribosomal protein S16 (BS17) [Bacillus halodurans]
 Identities = 62/90 (68%), Positives = 73/90 (80%)

Query: 1 MAVKIELRMGSKKGFYRINVDASRPDRGRFIRTVGTYNPLVAENQVTIKEERVELNL 60
 MAVKIEL RMGSKK FFYR+ VADSR+PRDGRFIE +GTYNEL +V +KE+R L+W+
 50 Sbjct: 1 MAVKIELRMGSKKAPFYRVVVDASRPDRGRFIRKIGTYNPLTPAKVLEKRLDWM 60

Query: 61 SKGAQPSDTVRNLLSKAGVNTKFDQKFSK 90
 KGA+PSDTVRNL SKAG+M K H+ K K
 Sbjct: 61 LKGAQPSDTVRNLPKAGLMEKLNKAKNEK 90

55

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2523> which encodes the amino acid sequence <SEQ ID 2524>. Analysis of this protein sequence reveals the following:

-918-

Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3654 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 86/90 (95%), Positives = 89/90 (98%)

Query: 1 MAVKIRLTRMGSKKKPFRINVDASRAPDRGRFIETVGTYNPLVAENQVTIKBERVLEWL 60

MAVKIRLTRMGSKKKPFRINVDASRAPDRGRFIETVGTYNPLVAENQ+TIKB+KVLKWL

15 Sbjct: 1 MAVKIRLTRMGSKKKPFRINVDASRAPDRGRFIETVGTYNPLVAENQVTIKBERVLEWL 60

Query: 61 SKGAQPSDITVRNLSKAGVMTKPHDQKFSK 90

SKGAQPSDITVRN+LSKAGVM KPHDQKFSK

Sbjct: 61 SKGAQPSDITVRNLSKAGVMAKPHDQKFSK 90

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 836

A DNA sequence (GBSx0888) was identified in *S. agalactiae* <SEQ ID 2525> which encodes the amino acid sequence <SEQ ID 2526>. Analysis of this protein sequence reveals the following:

25 Possible site: 35

>>> Seems to have no N-terminal signal sequence

30 INTEGRAL Likelihood = -11.09 Transmembrane 22 - 38 (16 - 42)

INTEGRAL Likelihood = -7.64 Transmembrane 382 - 398 (375 - 402)

INTEGRAL Likelihood = -7.59 Transmembrane 291 - 307 (284 - 317)

INTEGRAL Likelihood = -4.94 Transmembrane 340 - 356 (335 - 366)

----- Final Results -----

bacterial membrane --- Certainty=0.5437 (Affirmative) < succ>

35 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC24912 GB:AF012285 YknZ (Bacillus subtilis)

40 Identities = 161/417 (38%), Positives = 241/417 (57%), Gaps = 25/417 (5%)

Query: 1 NENWKFPALSSILGHKRAFLIMLGI IIGVAVSLINALGKGNKDSVINEITKSKQLQIY 60

+EN + ALSE+L HQR+ LIMLGI IIGV SV+++A+G+G + + I+ +++Y

Sbjct: 4 LENTRALSSVLAHKHRSILIMLGI IIGVSVI VVAVGQSGKMLKQSISSPONTVELY 63

45 Query: 61 YKTKEDQ-KNEIDNFGAQGFAPMGSDITNRKEPTIQBSWLKKIAKEVDGVSQYTVINQINAP 119

Y +++ + N A+ F + K K ++G+ + +

Sbjct: 64 YMFSDERLASNPAAAEFTPTINDI-----KGLGIGIKQVVAISTSEBK 109

50 Query: 120 VAYLEKIAKTVNITGIMRTYLGKPKFKISGRQFQEDYNQPSRVILLEKLAQR+LQITN 179

Y E++ + GIN Y+ + KI+SGR F + D+ +RV ++ +K+A+ LF

Sbjct: 110 ARYHEEETDAT-VNGINDGYMMNVLKISRGRTFTDNDPLAGNRVGIISQRMKELFKTK 168

Query: 180 EALANKVTVTKNSYLVGVGYSDPENGSLGYGSMEDGMALLINTQLASEPGAKEANIYF 239

+ L +VV + + ++GV +GL + + N + S PG + N+

55 Sbjct: 169 -SFLGEVWINGQPVETIIGVLKKV---TGLSFDLSMTYVFN-MKSSFGTSPDFENEL 223

Query: 240 HLNDVSGENRIGKEIKGRLLTDSHKADGYDNFDMTSIVKISINTQVIMNGVIGATAIS 299

+ GKE + + D +H + Y +M I I INT +IG+IA IS

60 Sbjct: 224 QVESADDIKSAGKRAQLVND-NHGTEDSYQVMNMBETAGIGKVTAIMTITIGTAGIS 282

Query: 300 LLVGGIGVMNIMLVSVTERTRKIGLRKALGATRRKILAQFLIESMVLITIGGLIGLLAY 359


```

LLVGGIGVNMIMLVSVIRREIG+RK+LGATR +IL OFILES+VLT++GGL+G+ + Y
Sbjct: 283 LLVGGIGVNMIMLVSVIRREIGIRKSLGATRGQILTQFLIESVLTLLGGHVGIGIGY 342

Query: 360 GGTMLIANAQDKITES-VSLNVAIGSLIFSAFIGLIIFGLLPANKAKSLNPIDALRYE 415
5      GG L++      PS +S V G ++PS IG+IFG+LPANKA+KL+PI+ALRYE
Sbjct: 343 GGAALVSAIAG--WPSLISWQVVOGSLVPSMLGIVIFGMLPANKAKLIDPILALRYE 397

```

There is also homology to SEQ ID 1350.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 837

A DNA sequence (GBSx0889) was identified in *S. agalactiae* <SEQ ID 2527> which encodes the amino acid sequence <SEQ ID 2528>. This protein is predicted to be ABC transporter (ATP-binding prot). Analysis of this protein sequence reveals the following:

```

15      Possible site: 52
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
20      bacterial cytoplasm --- Certainty=0.4080 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BA06841 GB:AP001517 ABC transporter (ATP-binding protein)
[Bacillus halodurans]
25      Identities = 131/218 (60%), Positives = 169/218 (77%)

Query: 8 LIRLEQIVKSYQNGDQKLQVLKNIDLTVYEGFLAIMGPSGSGKSTLNNIIGLLDSP 67
      +I+L ++ KS++ G + +++L IDL + G+FLAIMGPSGSGKSTLNNIIG LD PISG
30      Sbjct: 1 MIKLERVTKSFRVGTIEMVEILSAIDLEIASGDFLAIMGPSGSGKSTLNNIIGCLDQPTSG 60

Query: 68 DYSLANGRVERLSQTKLAQVRNKEIGFVPOQFLLSKLTALQNVLEPLIYAGVPPKKRN 127
      Y +GK + S+ ++A++RN+ IGFFVPOQ LL +L/TALQNVLEP++YAG+ K+R
35      Sbjct: 61 RYMFQKDLITNSQCEIAKIRNRHIGFVPOQHLLPR/TALQNVLEPMVYAGMKKERT 120

Query: 128 LAKQFLDKVELRERNNHILFTELSGGQKQKRVALARALVNSPSIILADEPTGALDTKTGEQI 187
      A L++V L ERN +LP LSGGQKQKRVALARALVNSPSIILADEPTGALDTKT E I
35      Sbjct: 121 RAHALERVGLAERMTYLPNSLGGQKQKRVALARALVNSPSIILADEPTGALDTKTSETI 180

40      Query: 188 MQFLTELNGEGKTIIMVTHPEIADYATRKIVIRDEGI 225
      M+ L LN EG TI +VTHPEIA+Y + + +RDG+I
      Sbjct: 181 MELLCSANNEGTTIALVTHPEIAEYTCQVFVVRDGI 218

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2529> which encodes the amino acid sequence <SEQ ID 2530>. Analysis of this protein sequence reveals the following:

```

45      Possible site: 52
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
50      bacterial cytoplasm --- Certainty=0.1739 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

55      Identities = 182/232 (78%), Positives = 207/232 (88%)

Query: 5 RKEILRLHIVKSYQNGDQKLQVLKNIDLTVYEGFLAIMGPSGSGKSTLNNIIGLLDSP 64

```

-920-

+K+L+L IVKSYQNKIDQ L+VLK I+LTVYBSEPLADMGPSGSGKSTLANNIIGLLD P
 Sbjct: 5 KQIAQLSNIVKSYQNCQVLKVLKGINLTVYBSEPLADMGPSGSGKSTLANNIIGLLD P 64
 Query: 65 TSGDYSLNGKRVSELSCTKLAQVNRKRGVFPVQFFLLSKLTALQNVPLIYAGVPPKK 124
 5 TSGDY+L+ ++B L+ +LA+VRN EIGFVQQFFLL+KLTLQNVPLIYAGV K
 Sbjct: 65 TSGDYTLHTTKIEILNDRBLAKVRNDEIGFVQQFFLLKLTLQNVPLIYAGVNVSK 124
 Query: 125 RKNLAKQFLDKVELREFRNHLPTLSGQKQKVAIARALVNSPSIIADEPTGALDTTKG 184
 10 R+ AQLFL+KV L R+ HLP+ELSGGQKQKVAIARALVNSPSIIADEPTGALDTTKG
 Sbjct: 125 RRECAQFLKVLKGLRRRIKHLSELSGGQKQKVAIARALVNSPSIIADEPTGALDTTKG 184
 Query: 185 EQIMQFLTELMOEGKTIIMVTHEPEIADYATRKIVIRDEITADTDSIRID 236
 +QIM+ LTELW+EGKTIIMVTHEPEIAD+ATRKI+IRDG+IT DTT S+ ID
 Sbjct: 185 QCIMELLTELNKGKTIIMVTHEPEIADPATRKIIIRDGITTTTASVVID 236
 15

Based on this analysis, it was predicted that these proteins could be useful antigens for vaccines or diagnostics.

Example 838

A DNA sequence (GBSx0890) was identified in *S. galactiae* <SEQ ID 2531> which encodes the amino acid sequence <SEQ ID 2532>. This protein is predicted to be ATP-binding cassette transporter-like protein. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -8.97 Transmembrane 17 - 33 (13 - 39)
 25 ----- Final Results -----
 bacterial membrane --- Certainty=0.4588 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9965> which encodes amino acid sequence <SEQ ID 9966> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA24909 GB:AF012285 YknX [Bacillus subtilis]
 35 Identities = 104/391 (26%), Positives = 182/391 (45%), Gaps = 21/391 (5%)
 Query: 13 KKGATISGLVALIVVIGGFLNVQSQPNKSAVKTYKRVFNVRGSGVSSSTLLTGKAKAQ 72
 KK I G+V + + +G ++ + P + + +V E +SS+ ++ G K +
 Sbjct: 2 KKVWIGIGIAIVIALPVGINDIYRSAAPTSGSAGEVQTGVSVERNEISSTVMVPGTKLPSN 61
 40 Query: 73 EQYVYFDANKENRATVTVKVGDKITAGQQLVQYDITTAQAAYDTANRQLKVARQINNLIK 132
 EQYV++A+K + VK GDK+ G LV Y T Q + + GL + ++ +
 Sbjct: 62 EQYVYFADKQTLIEDIKVKGDKVKKGDTALVTV--TNEQLSLEKQKQQLTSESNRIQIQD 119
 45 Query: 133 TTGSLPAMESDQSSSSSGCGGTOSTSGATNRLQNYQSCANASYNQQLQINDAYADAQ 192
 L A++S ++ G+ + R + Q + +L Q
 Sbjct: 120 IQEKLKALDSKERLEKQVGKKRAEQKESRETELQMKKTAEI-----ELKQTELQRF 173
 Query: 193 ARVNAKQAKALNDITVITSDVSGTVVEVNSDIDPASKTSQV---LVHVATBGLQVQGTMS 249
 + N+ ++D + S++ GTV+ VN + ASK S + ++H+ L V G +SE
 50 Sbjct: 174 SLANR----VSDLVKSBIEGTIVISVWQ--EASKKSDIQEPVHIGNPKDLVSGGLSE 227
 Query: 250 YDLANVKKDQAVKICKSKVPYDKKWEKGISYISNYPEARNNDSINNGSNAVNYKYKVDIT 309
 YD VKK Q V + S V K W+G+S + P+ + + + AV Y +V I
 55 Sbjct: 228 YDYLKVKKGKGVITSDVIGQKTKWGTYSAGVLVDP--QOBSAAAGGTGQAVYPLQVQKIK 286
 Query: 310 SPLDALKQGFVSVEV-VNGDKHLIVPTSSVINKNKHFPVWYVNSNRKISKVEVKIGKA 368
 L K GF + + + K +P+ +V +D+++V+ D K +V+VKIG+
 Sbjct: 287 GNLPEKGPGEKFINIRITDKRKANTLPSKAVKGGDQYVYTVKDG--KAKRVVDVKGQ 344

Query: 369 DAKTQRIILSGLEAGQIVVNFSTKPKDQKI 399

EI GL V+ NPS DG ++

5 Sbjct: 345 TDDLTEIKRGLTQDDQVILNPSDQVTDGMEV 375

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2533> which encodes the amino acid sequence <SEQ ID 2534>. Analysis of this protein sequence reveals the following:

Possible site: 42

>>> Seems to have an uncleavable N-term signal seq

10 INTEGRAL Likelihood = -9.61 Transmembrane 15 - 31 (11 - 36)

----- Final Results -----

bacterial membrane --- Certainty=0.4843 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

15 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AC24909 GB:AF012285 YmcX [Bacillus subtilis]

Identities = 103/380 (27%), Positives = 180/380 (47%), Gaps = 21/380 (5%)

20 Query: 16 ITASVITLVLIITIGVILMKQQRNLTADIAKEPYSTVSTEGSIASSTLISGTVKALSEE 75

I + + +V + + GI + + + T + A + T SV E I + + + + GT + K + E +

Sbjct: 6 IGIGIAIVIALFVGINIVRSAPT--SGSAGKVCVTGSVZENISSTVWVPGTILKFSNEQ 63

25 Query: 76 YIYFDANKNDATVTVKVGDDVTQGGQLVQYNTTITTAQSAIVTAVRSLNKIGRIINHLATY 135

Y + + + A + K G + VK GD + V + G LV Y T Q + + + N + + + N L +

Sbjct: 64 YVYFADKGTLEDIKVKESDVKHGTALVTY--TKQLSLE---KEQNLTSENRQLID 118

30 Query: 136 GVPVAVSTETNRDEATGEETTTTVPQSAQ--QNAVYKQQLDANDAYADAQAEVNTAQIA-- 192

+ + + E E + + Q + + + Q + Q Q E + + + A +

Sbjct: 119 QIQEKALKALDSKRELEKQVKKKEAKQIESRETELQMQKKTARTELKQTELQKQSLANK 178

35 Query: 193 LNDTVIISVSGTVVVRND--IDPSKNSQTLVHVATEGQLQVGTILTEYDLANKVVGQS 251

+ + D V S + GTV + VN + S + + H + L V G L + EYD VK GQ

Sbjct: 179 VSDLEVKSEIBGTIVISVQEAASKSDIQEVPVHIGNPDLVVGSKLSEYDPLAVKQCK 238

40 Query: 252 VKIKSKVYSNQEWTKGISVSVNYPTESNAGSTTPAGSTGASSTGATYDYKIDILSPINQ 311

V + S V + W G + S V P + + + G + Y + + I L +

Sbjct: 239 VTLTSDVIQGTWKGTVSAVGLVFDQGES-----AAQGTGEAVQYPLQVKIKNLP 291

45 Query: 312 LKQGTIVSVVVENAKQA-LVPLTAVIKKDKKHVWYTDATGAKAKVEVTLGNADAQQ 370

K GF + + + + A + P AV K + D + + YV + T D GKAK + V + + G

Sbjct: 292 GKPGPKPINNIETDKRKNLTLPKAVKKEDDQYVYTVKD--GKAKRVDKVIGSVTDLIT 349

50 Query: 371 EIKHGAAGDVIANDPKNI 390

EI + G + D VI NP +

Sbjct: 350 EIKRGLTQDDQVILNPSDQV 369

An alignment of the GAS and GBS proteins is shown below.

50 Identities = 234/421 (55%), Positives = 301/421 (70%), Gaps = 19/421 (4%)

Query: 3 MSKRNLGISKKGAIISGLSVALIVVIGGF-LAVQSQPNKSA--VKINIKVFWNRBSVS 59

MSKR + I + K + I + + L + + I G LM Q + + A K Y + V EGS + +

55 Sbjct: 1 MSKRKIKIKITTKTLITASVITLVLIITIGVILMKQQRNLTADIAKEPYSTVSTEGSIA 60

Query: 60 SSTLLTGKAKAMQSVYFDANKGNRAVTVVKVGDKITAGQQLVQYNTTAAQAAYDTANR 119

SSTLL + G KA E + I + YFDANKGN ATTVVKVG + + T GQQLVQY + TTAQ + AYDTA R

Sbjct: 61 SSTLLSGTVKALSSEIYFDANKNDATVTVKVGDDVTQGGQLVQYNTTAAQSAIVTAVR 120

60 Query: 120 QLNKVARQINMLKTTGSLFAMESSDQSSSSSQGQSTGSGATNRLQNYQSQNNASVQ 179

LNK + RQIN + LKT G + FA + S + + + G + T + + + Q NAA + Q

Sbjct: 121 SLNKIGRIINHLATYG-VPAV-STETNRDEATGEETTTTVPQS-----AQGNANYQ 170

Query: 180 QLQDLNDAYADAQAEVNTAQKALNDATVITSVSGTVVEVNSDIDPASKTSQVLVHVATEG 239

-922-

QIQQLNDAYADAQAEVNIKAQ ALMDTV+ S VSGTVVEVN+DIDP+SK SQ LVHVATEG
 Sbjct: 171 QIQQLNDAYADAQAEVNIKAQIALMDTVVSSVSGTVVEVNDIDPSSKNSQTLVHVATEG 230
 Query: 240 KLQVQGTMSFYDLANVKKDQAVKI KSKVYDKHWEKGKISVSNYP+BAEANN-----NDS 293
 +IQV+GT++HYDLANVK Q+VKIKSKVY ++EW GKISY+SNYP R+ A + +
 Sbjct: 231 KLQVQGTTLTRYDLANVKVQGSVKIKSKVYSNQEWTKISVSNYPSTESNAGSTTPAGSTG 290
 Query: 294 NNGSSAVNYKYKVDITSLDALKQGPVTSVEVVNGDKHLIVPTSSVINKNKHFWVYND 353
 S+ Y YK+DI SPL+ LKQGPVTSVEVVN K +VP ++VI KD KH+VW Y+D
 Sbjct: 291 AGSSTGATYDYKIDIISPLNQLKQGPVTSVEVVNKAQALVPLTAVIKDKKHVWYVTD 350
 Query: 354 SNRKISKVEVKIGKADAKTQEI LSLGAKGQIVVTNPSTKTFDQKIDNIESIDIANSKKSE 414
 + K KVEV +G ADA+ QEI G+ G IV+ NP K K +K++ + SI N+ + +
 Sbjct: 351 ATGAKKVEVILGNADACQQRHKGVAWGDIVIANPDKNKKPKKLEGVISIGINIKPEED 411

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 839

A DNA sequence (GBSx0891) was identified in *S. agalactiae* <SEQ ID 2535> which encodes the amino acid sequence <SEQ ID 2536>. Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1832 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 840

A DNA sequence (GBSx0892) was identified in *S. agalactiae* <SEQ ID 2537> which encodes the amino acid sequence <SEQ ID 2538>. This protein is predicted to be carbamoyl-phosphate synthase, pyrimidine-specific, large chain, putati. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -1.70 Transmembrane 486 - 502 (486 - 502)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1680 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA91005 GB:Z54240 carbamoyl-phosphate synthase [Lactobacillus
 plantarum]
 Identities = 117/417 (28%), Positives = 205/417 (49%), Gaps = 37/417 (8%)
 Query: 122 FVQVDCILVWRDLSANCLYVDLEYIES-NKTKESLAIVPSQTLSDAARQTIRIVAFVVC 180
 + +++ VNRD+ +N + V + +E + L TG S+ P QTL+D Q +RD A +
 Sbjct: 213 YKEIEFVNRDAADNANVVCNMFENFPGVHTGDSIVYAPVQTLADREVQLLEDAALKITI 272

-923-

Query: 181 RKANLIQVCFPSFLIDIASLDYHIIISLSSGLSHQSILFTITITYPVLEIATKLVGTYPS 240
 R I G C +D NS +Y+II ++ +S S L T YP+ ++A E+ VG
 Sbjct: 273 RALKIBOGCNVQLADLPNSFNYYILEVNPVRVSRSALASKATGYPIARMMAKIAVLGLD 332

5 Query: 241 QLKHSYYPNTSAFLRQLOLVATV--SFSFEKVDY-----IFFARNIQOL 283
 ++K+ T A RP LDYV +F+K + + RNIE+
 Sbjct: 333 EIKNFVITGTYYAFSPALDYVVKIIPWFDFKFTHADRLTQMKATGEVMAIGRIERA 392

10 Query: 284 FINLLEASS----HDHFFFLDISSEDLMFALIQKKNRGLAYLEAFRRGPDLYGLSSVT 339
 L + + H L + ++ L LI ++RL YL EA RRG+ +L+ +T
 Sbjct: 393 TLKAVRSLGIVHVESLRSVDDVLSDKLIHQDORLFYTBARIHQYQIDELAEIT 452

Query: 340 KINFFYLKCLHIVELYENLANSQYNVDIYKRAKRYGSDDYIASSWQISLIDMLBYRKK 399
 KIN F+LKK LHI+E+ + L +++ AKR GF+D +A W ++ + ++R
 15 Sbjct: 453 KINFFYLKCLHILEQALRTHTDIDIELTVAKRNGFADQIVADYWHETIDQVRDFLA 512

Query: 400 HSAVFLKQVQSSGVLTGHQIQYFRSYDWSYDISSGQKALIM-----VDKGY 449
 H +AFV K V+ +G Y+ +Y++ ++ I + L++ V+ Y
 20 Sbjct: 513 HGAFAVYRNVDTNCFGEFASFTPYTYGYEFENESIVTKRPSVLVIGSGPIRGQVGFEDY 572

Query: 450 SLVKINLEIKQIKOTHELELLIVTNPQLLEIQLNDTS--IIPDTIGIETILTIMGIE 504
 + V +K I++ E +I+ + P + S + F+ + IB +L ++ +E+
 25 Sbjct: 573 ATV--HSVKAIQAGYBALINNSNFTSVSTDSVSKLYFELTIEDVINLELAK 626

- 25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 841

A DNA sequence (GBSx0893) was identified in *S. agalactiae* <SEQ ID 2539> which encodes the amino acid sequence <SEQ ID 2540>. This protein is predicted to be carbamoyl phosphate synthetase small subunit (carA). Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.2709 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:CAB89872 GB:A132624 carbamoyl phosphate synthetase small
 subunit [Lactococcus lactis]
 Identities = 188/352 (53%), Positives = 265/352 (74%)

45 Query: 1 MAKLLILEDDTVPBGLPGSSLDVTGELVFTGNTGYQELITNPEHNGKILVFTSPLIG 60
 M+K+LLILEDDT+PEG + G+LDVIGELV TG TGYQ IT+ S+NG+IL FT P++G
 Sbjct: 1 MSKRLLILEDDTIPBGEALGANLVGTGELVFNVTGNTGYQSSITDQSYNGQILFTPTVIVG 60

Query: 61 NYGIHRSYSEAIIPYCLGVVVAEYSRVSSDTSKNNLDEFLKMKKVPAMSGVDTRVLMQ 120
 NYG++R E+I PTC VVV E +R S+ +M+ DEFLK K +P ++GVDR + ++
 50 Sbjct: 61 NYGVNRDDYSIHPCTKAVVVEAARRPSNWMQMSDFSLKKNIPGIVGVDTATIKI 120

Query: 121 IKEKGFKVKTAEAGDVLSHLOQLIATVLPNNVSGVSTKTATYPSASGRNIVVLDLGL 180
 ++E G +KA+L +A D + H QL ATVLPN VR ST TAYPS +GR +VV+DFGL
 Sbjct: 121 VREHGTMKASLVQARDEVDMQMSQLQATVLPNQVETSSSTATYPSPTGRKVVVDLGL 180

55 Query: 181 KHSILRELKSKQCDVTPVYNTSLEGIKNLYPBGIIISNGPGNPEKIQSLINTIKELQS 240
 KHSILRELKSKR+C+TV+FYNTS + I + P+G+L+NGPG+P + E + IRE+Q
 Sbjct: 181 KHSILRELKSKRECNLTVPVYNTSAKRIEMEPDGVMLTNGPGDPTVPEAEMIKESVGK 240

60 Query: 241 VFMLIGLGHQLAMANGAEIMRLPVAKEGPNYPMDIATCRLETVSQNEFTVNRLLLP 300
 +P+ GI LQHQL ++ANGA ++ +G N+ +R++ATGR++ SQ + + V+ RLP

-924-

Subjct: 241 IPIPGICLGHQLFSLANGAITYKMKPGHRRGZNAHAREVATGRIDPTSQNHGYAVSSSENL 300

Query: 301 HDLVLVTHGINDQRIVALRHRSFPVMSVQFPEAAGPHDVTYFDFRFLGEMI 352
DL+ATH +ND + +RH+ PD SVQF+P+AARGPHD +Y ID+P++++

5 Subjct: 301 EDLMITHVKINDNSVSGVHHKYFEAPSVQFHPDAARGPHDASYLDDPMOLM 352

There is also homology to SEQ ID 2030.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

10 Example 842

A DNA sequence (GBSx0894) was identified in *S.agalactiae* <SEQ ID 2541> which encodes the amino acid sequence <SEQ ID 2542>. Analysis of this protein sequence reveals the following:

Possible site: 57
>>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3646(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20 A related GBS nucleic acid sequence <SEQ ID 9967> which encodes amino acid sequence <SEQ ID 9968> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:CAB89869 GB:AJ132624 pyrimidine regulatory protein [Lactococcus
lactis]
Identities = 127/169 (75%), Positives = 147/169 (86%)

Query: 13 MKRKEIIDVTMKRAITRITYEIIERNKLNIVLAGIKTRGVFIARQIERLQKLENLD 72
M RKEIID++TMKRAITRITYEIIERNK LD +VL GIKTRGV+LAKRIQERL+QLE L+

30 Subjct: 1 MARKKEIIDVTMKRAITRITYEIIERNKELKIVLIGIKTRGVFLAKRIQERLQKLE 60

Query: 73 IPFGELDTKPPRDDKIVZVDVTMPVDITDKDIIIDVLYTGTIRAAIDNLVSLORPS 132
IP QELDT+PPRDD + + DIT + +DIT KD+IL+DDVLYTGTIRAAID +V LGRP+

35 Subjct: 61 IPFGELDRPFRDDKQAGEDTTEIDIDITGKDVILDDVLYTGTIRAAIDGTVKLGRPA 120

Query: 133 RVSLAVLIDRGHRELPIRADYVGKNIPTSQFEELVAVMEHDGVDVRSV 181
RV LAVL+DRGHRELPIRADYVGKNIPT EET+V++ EHDG D + I

Subjct: 121 RVQLAVLVDRGHRELPIRADYVGKNIPTGHDEELVQVSEHDGNDLSILI 169

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2543> which encodes the amino acid sequence <SEQ ID 2544>. Analysis of this protein sequence reveals the following:

Possible site: 45
>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3870(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 An alignment of the GAS and GBS proteins is shown below.

Identities = 147/171 (85%), Positives = 158/171 (91%)

Query: 13 MKRKEIIDVTMKRAITRITYEIIERNKLNIVLAGIKTRGVFIARQIERLQKLENLD 72
MK KEI+DDVTMKRAITRITYEIIERNK LKN+VLAGIKTRGVFLA+RIQERL QLE LD

55 Subjct: 1 MKRKEI+DDVTMKRAITRITYEIIERNKLNIVLAGIKTRGVFIARRIQERLQKLELD 60

-925-

Query: 73 IPVGELOFKPFRDDMKVEVDITTFMEVDITDKDIIIDVLYTGRTIRAAIDNLVSLGRPS 132
 +P+GELO KPFRDDMK+VE DTF M VDT KD+IIIDVLYTGRTIRAAIDNLVSLGR+
 Sbjct: 61 LPIGELODKPFRDDMRVBEOTLMSVDITGKDVLIIDVLYTGRTIRAAIDNLVSLGRE 120

5 Query: 133 RVSLAVLIDRGHRELPIRADYVGKNIPTSQFEELVEMBEDQYDRVSIID 103
 RVSLAVL+DRGHRELPIRADYVGKNIPTS RRI+VEV+E DG DRVSIID
 Sbjct: 121 RVSLAVLIDRGHRELPIRADYVGKNIPTSSVEBIVVEVVDGDRVSIID 171

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 843

A DNA sequence (GBSx0895) was identified in *S.agalactiae* <SEQ ID 2545> which encodes the amino acid sequence <SEQ ID 2546> (rhuD). Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.0687 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9969> which encodes amino acid sequence <SEQ ID 9970> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06261 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 178/290 (61%), Positives = 216/290 (74%), Gaps = 2/290 (0%)

Query: 17 GVRLDKAL-ADNSBLGRSQANEIKGGIVLVNGQVKAICYTVQEGDRITFDIPKEVLDY 75
 G R+DK L A E SR+Q + IK G VL+NG+ K+ Y V+ GD + +P+ EVL+
 Sbjct: 15 GERIDKFLTAQGEBSRTQVQWIKDGHVLINGRTIKSNYKVTGDTLFLVEPEPRVLEV 74

Query: 76 QAENIPLDIIYQDDVAVVVKPQGMVHPSAGHSGGLVNALMYHIDLSSINGVVRPGI 135
 ENIP+IIY+D+DVAVVVKP+GMVHP+ GH++GTLVNALMYR DLSSINGVVRPGI
 Sbjct: 75 VPENIPRIIYSESDVAVVVKPRGMVHPAGPHTTGTLVNALMYHCHDLSSINGVVRPGI 134

Query: 136 VHRIDKDTSGELMVARNDRAHQVLAELKDKKSLRKYLAIVHGNLMDRGVIEAPIGRSD 195
 VHRIDKDTSGELM+AKNDRAH+ L +LK K + R Y AIVHGN+P+D G I+APIGR
 Sbjct: 135 VHRIDKDTSGELMIARNDRAHESLVNQLKAKTTERVYQAIVHGNIPHDHOTIDAPIGRDK 194

Query: 196 KDRKKQAVTAK-GKPAITRFHVLERPQDTYLVELSLETGRTHQIRVHMAYIGHFLAGDFV 254
 DR+ VT + + A+T F VLERPGD+T VE LETGRTHQIRVH YIG FLAGDF
 Sbjct: 195 VDRQMTVTEENSRDAVTHFTVLERPGDPTFVCEQLETGRTHQIRVHFKYIGFLAGDFK 254

Query: 255 YGPKKTLGGKQQLHAQTLGFTPHSPNGNLIFFSVEVPEIPQITILEKIRN 304
 YGP+KTL GQ LHAQ LGF HP GE + F VE+PE + + +L+ N
 Sbjct: 255 YGPKKTLSDGQALHAQKLGFSEHPTGEFMRFKVEMPERMKLIRQLQNN 304

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2547> which encodes the amino acid sequence <SEQ ID 2548>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.2455 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-926-

Identities = 239/295 (81%), Positives = 265/295 (89%)

Query: 9 MEITIKIAGVRLDKALADNSLSRSQANSEIKRGIVLVNQGVKKAKYTVQEGDRITFDIP 68
 MEI + +G RLDKALAD S LSR QAN++IK+G+VLVNGQ KKAKYTVQ GD I F++P
 5 Sbjct: 1 MEINVTISQORLDKALADLSPLSRGQANDQIKQGLVLVNGQKKAKYTVQAGDVICFELP 60

Query: 69 KEEVLVDYQANIFLDIYQDDVAVVNFQGMVVHPSAGHSSGTLVNALMYHIKDLSSIN 128
 KEEVL+YQA+NIFLDIY+DD +A++NKPQGMVVHPSAGH SGT+VNALMYHIKDLSSIN
 10 Sbjct: 61 KEEVLEYQAQNIPLDIYEDDVALAINPKQGMVVHPSAGHPSGT+VNALMYHIKDLSSIN 120

Query: 129 GVVPRGIVHERIDKDTSGILLMVAKNDRHQAQVLAELKDKKSLRKYLAIVHGNLNDRGVIE 188
 GVVPRGIVHERIDKDTSGILLMVAK D ARQ LAELK KSLRKYLAIVHGNLNDRG+IE
 15 Sbjct: 121 GVVPRGIVHERIDKDTSGILLMVAKTAARQAQVLAELKAKKSLRKYLAIVHGNLNDRGVIE 180

Query: 189 APIGHSKDKRKQAVTAKGKPAITRPHVLERFGDYTVLSELTGRTHQIRVHMAYIGHP 248
 APIGRS+KDKRKQAVTAKG A+TRF VLERFGDY+LVEL LETGRTHQIRVHMAYIGHP
 15 Sbjct: 181 APIGRSEKDKRKQAVTAKGKPAITRPHVLERFGDYTVLSELTGRTHQIRVHMAYIGHP 240

Query: 249 LAGDPLVVGPRKTLGGKQFLHAQTLGTHPSKGNLIFSVEVPEIQTTLKLRK 303
 +AGDP+VGRKTL G GQFLHA+TLG THP G+ +IF+VE PEIQT TLKLRK
 20 Sbjct: 241 VAGDPLVVGPRKTLGGKQFLHAQTLGTHPSKGNLIFSVEVPEIQTTLKLRK 295

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 844

A DNA sequence (GBSx0896) was identified in *S.agalactiae* <SEQ ID 2549> which encodes the amino acid sequence <SEQ ID 2550>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence
 30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0496 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 35

The protein has homology with the following sequences in the GENPEPT database.

>GF:AADS3064 GB:AF163833 CpeY [Streptococcus agalactiae]
 Identities = 105/297 (35%), Positives = 163/297 (54%), Gaps = 4/297 (1%)

Query: 1 MNIQQLRYVVAIANSOTPREAAALFVSPQSLSVAVRDLSTELGPQIPTRTTGAVLGNQ 60
 M IQQT+YV+ I +G+ EAA +L+++QPSLS AVR+LSTE+G QIF R G LT
 Sbjct: 1 MRIQQLQYVIVKIVETGSMNEAAKQLTITQPSLSNAVRNLTETMGQIFIRNPKGILPLTKD 60

Query: 61 GMTFVENALEVVKSFDSFEKQSQSEATEQRFSLASQHYDLEPLPLATAPSKCNPNFSY-F 119
 GM F A +++ E+++ + + PS++SQHY F+ A D Y
 45 Sbjct: 61 GMEFLSVARQLLEQTALLERYKQDNTSRFLPSVSSQHYAFVUNAFVAFVNGTMDQYEL 120

Query: 120 RIFESTTIRILDEVAQNSIEGLIYNSQNKGLLQRLDKLGEVLEPLPFKTHIYAGK 179
 + E+ T I+D+V SEIG++NNS N+ L+ D L L L HI++ K
 50 Sbjct: 121 FLETRTWELIDOVKNFRSEIGVLFNSYNRDLVTLKFDONSLIATLFTTTPHIFVSKS 180

Query: 180 HPIASKTSLIMTDLSEGLPTVRFTQDRDYRYSENFEVLDSSVTYVNRDRATNGILER 239
 +PIA++ L M DLE P + + Q + Y+SE + + + V+DRATL +
 Sbjct: 181 NPLANRKLKSMKDLIEDYPLSYDQGLANSFPSEEMMSQIPHEKSIIVSDRATLNLIMIG 240

Query: 240 TQAYATGSGFLDSRVNG--ITVPLEHLDNQMIYIKRDRNLSQMAKPKVAVMSE 294
 Y +G L+S+ +NG I IPL+ ++YI+ NLS+M KP+ + E
 Sbjct: 241 LDGYTVATGILNSK-LNGDRIVAPLDVDDVIDIYIRIDKANLKRQKQFDIYILE 296

60 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2551> which encodes the amino acid sequence <SEQ ID 2552>. Analysis of this protein sequence reveals the following:

-927-

Possible site: 13
>>> Seems to have no N-terminal signal sequence

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1252 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

10 Identities = 217/296 (73%), Positives = 253/296 (85%)

Query: 1  MNIQQLRYVVAIANS+GTFRFAAKLFSVQPSLSVAVRDLETLGFCQIFRTTTGAVLNIQ 60
      MNIQQLRYVVAIAN+GTFRFAA+KLFSVQPSLSV+++DLE ELGFCQIP RTT+G VLI+Q
Sbjct: 1  MNIQQLRYVVAIANNKTFRAASKLFSVQPSLSVSIKDLFAELGFCQIPNRTTSGTIVLSQ 60

15 Query: 61  GMTTFYENALEVVKSFDSFENQPSQSEATQBEFSIASQHYDFLPLLTAFSKCONDNSFYFR 120
      G+ FYE ALEVVKSFDSFEK FSQ++ + EFSIASQHYDFLPLLTAFS+ D FR
Sbjct: 61  GLVIFYKALEVVKSFDSFEKTFQSADLDQNEFSIASQHYDFLPLLTAFSQQYDGHVRVFR 120

20 Query: 121 IFESTTIRILDEVAQGNSEIGIYYINSQNKGLLQRLDKLGLFVELIPFKTHIYLGKH 180
      IFESTTII+ILDEVAQGNSEIGIYY-N N+KGL QR+DKLGLS+V LIFF THIYL K H
Sbjct: 121 IFESTTIQILDEVAQGNSEIGIYYINVDNQKGLQFORMDKLGLFVSLIPPTTHIYLSKTH 180

25 Query: 181 PLAKTSLIMTDLGLELPTVRFTQDRDDYRYISENFVVLDSVTVYNDVDRATLNGILERT 240
      PLA++ +L + D++GLP VRFTQ+RD+Y YYSENFV+ + YNV+DRATLNGILERT
Sbjct: 181 PLANRRELYNDIQGLFAVRFTQERDEYLYISENFVVTSECPRIYNVSDRATLNGILERT 240

Query: 241 QAVATSGFLDSRVNGIIVIPLEDLNLMYIKRKDRNLQMAKLFVAVMVEYF 296
      A+ATGSGFLD RSVNGI VIEL DH+DNQMIY+KRKD+NLS FV ++++YF
30 Sbjct: 241 NAFATSGFLDHRSVNGIKVILPLADHIDMIMYIVKRKDRNLVAGATFVTILKDYF 296

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 845

35 A DNA sequence (GBSx0897) was identified in *S. agalactiae* <SEQ ID 2553> which encodes the amino acid sequence <SEQ ID 2554>. This protein is predicted to be 50S ribosomal protein L27 (rpmA). Analysis of this protein sequence reveals the following:

```

Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0976 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14754 GB:Z99118 ribosomal protein L27 (BL24) [Bacillus subtilis]
Identities = 70/90 (77%), Positives = 80/90 (88%)

```

50 Query: 8  NLQLFAHKKGCGSTNGRDSQAKRLGAKAAGCTVSGGSLYVQRGTHIYPGANVGRGGD 67
      +LQ FA KKG GST NGRDS+AKRLGAK ADQ V+GGSLYVRQGT IYGP NVGRGGD
Sbjct: 5  DLQFPASKKGVGSTKNGRDSAKRLGAKRADQPVVGGSLYVRQGSTKIYPGENVGRGGD 64

Query: 68  DTLFAKVGIVRFRERGRDRKKQSVYPIAK 97
      DTLFAK+G V+FER GRD+K+VSVYF+A+
55 Sbjct: 65 DTLFAKIDGTVKFERPGRDRKKQSVYFVAQ 94

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2555> which encodes the amino acid sequence <SEQ ID 2556>. Analysis of this protein sequence reveals the following:

-928-

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

5 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0976 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 95/97 (97%), Positives = 96/97 (98%)
 Query: 1 MLKPKLANLQLFAHKKGSGSTNGRDSQAKRLGAKAADGQTVSGGSILYRQRGTHIYPGA 60
 MLKPKLANLQLFAHKKGSGSTNGRDSQAKRLGAKAADGQTVSGGSILYRQRGTHIYPG
 Sbjct: 1 MLKPKLANLQLFAHKKGSGSTNGRDSQAKRLGAKAADGQTVSGGSILYRQRGTHIYPGV 60
 Query: 61 NVGRGGDDTLFAKVBGVVRFERKNGRDKKQVSVYPIAK 97
 NVGRGGDDTLFAKVBGVVRFERKNGRDKKQVSVYP+AK
 Sbjct: 61 NVGRGGDDTLFAKVBGVVRFERKNGRDKKQVSVYPAK 97

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 846

A DNA sequence (GBSx0898) was identified in *S. agalactiae* <SEQ ID 2557> which encodes the amino acid sequence <SEQ ID 2558>. Analysis of this protein sequence reveals the following:

25 Possible site: 25
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.75 Transmembrane 32 - 48 (32 - 48)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1298 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:BA06729 GB:AP001517 unknown conserved protein in B. subtilis
 [Bacillus halodurans]
 Identities = 33/107 (30%), Positives = 63/107 (58%), Gaps = 4/107 (3%)
 Query: 1 MIKATPTRNQSGYLYSABISGHAGSGEYGFVICAAVSTLSINFINLEALITQCAQLII 60
 MI F RN+ + S +SGHA +G YG D++CA S +++ +N++ AL CQ +L+
 Sbjct: 1 MIDVVFERNKNDIVSFTHSGHADGPGYQQLVCAGASVALGTVNAIIAL--CQVELVT 58
 Query: 61 N-DVEGGYMKIDL-SSIPQHKEDKVLQLLFESVILLGMNLSKDSSEFV 105
 + EGG+++ + + + + +KVQLL E + + + + + E +
 Sbjct: 59 EMENEGGLRCRVFNDLESTTFERKVLQLLFGMNLISLQSLAESYGRHI 105

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2559> which encodes the amino acid sequence <SEQ ID 2560>. Analysis of this protein sequence reveals the following:

50 Possible site: 52
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.59 Transmembrane 32 - 48 (32 - 48)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1235 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

-929-

>GP:BA06729 GB:AP001517 unknown conserved protein in *B. subtilis*
 [Bacillus halodurans]
 Identities = 33/109 (30%), Positives = 60/109 (54%), Gaps = 4/109 (3%)

5 Query: 1 MIKALFTRQKQGLSSVTLTGHAGSGKHGFDIVCASVSTLAINFVNSLEVLADQCALVDL 60
 MI +F R K + S T+GHA +G +G D+VCA S +A+ VN+ L + + + +
 Sbjct: 1 MIDVVFERNKQNDIVSFMTSGHADAGPYQDVLVCGASAVALGTVNAIILALCQVELVTEN 60

10 Query: 61 NDVEGGYMAITIP---PHDKKEEVQLLFESFLGNTLSKDSKFPVNTQ 106
 + EG3+ + P E+VQLL E + + S+A+ + + +
 Sbjct: 61 EN-EGGFLRCRVPNDLEETTFEKKVQLLBCMNISLQSIABSYGERHQIE 108

An alignment of the GAS and GBS proteins is shown below.

Identities = 67/110 (60%), Positives = 90/110 (80%), Gaps = 2/110 (1%)

15 Query: 1 MIKALFTRNQSGYLSAEISGHAGSGSYGFDVCAAVSTLSINFINSLEALTCQALII 60
 MIKA FTR +G L S +G HAGSG+GFD+Ch+VSTL+INF+NSLE L CQA + +
 Sbjct: 1 MIKALFTRQKQGLSSVTLTGHAGSGKHGFDIVCASVSTLAINFVNSLEVLADQCALVDL 60

20 Query: 61 NDVEGGYMKIDLSSIPCHKEDKVLLEFESYLLGNTLSKDSSEFVTVVM 110
 NDVEGGY I + P +++VQLLFES+LLGNT+L+KDS+FV+T V+
 Sbjct: 61 NDVEGGYMAITIP---PHDKKEEVQLLFESFLGNTLSKDSKFPVNTQVI 108

SEQ ID 2558 (GBS433) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 78 (lane 4; MW 16kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 8; MW 41kDa).

GBS433-GST was purified as shown in Figure 223, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

30 Example 847

A DNA sequence (GBSx0899) was identified in *S. agalactiae* <SEQ ID 2561> which encodes the amino acid sequence <SEQ ID 2562>. This protein is predicted to be ribosomal protein L21 (rplU). Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2972 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 40 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CA14756 GB:Z99118 ribosomal protein L21 (BL20) [Bacillus subtilis]
 Identities = 67/101 (66%), Positives = 78/101 (76%)

45 Query: 4 YALIKTGKQKVKVSVGQAIYVEKLDVVGAEVTFNEVVLVGGSETTKVGTPTVVBGATVVG 63
 YALIKTGKQK+KVE GQ +Y+EKL RAG VTF +V+ VGG+ KVG P VEGATV
 Sbjct: 2 YALIKTGKQKIKVBEQGTVYIEKLAARAGETVTFEDLVFGQDNVKNPTVVBGATVTAK 61

50 Query: 64 VEKQCKQKVKVSYKYKPKKGSHRQGHRCQPTTKVIVINAINA 104
 VEKQG+ KK+ ++YKPKK H+KQGHRCQPTTKV I INA
 Sbjct: 62 VEKQGRAKKLTVERKYKPKKQNVHKQGHRCQPTTKVTIEKINA 102

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2563> which encodes the amino acid sequence <SEQ ID 2564>. Analysis of this protein sequence reveals the following:

-930-

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 5 bacterial cytoplasm --- Certainty=0.3026(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- 10 Identities = 97/104 (93%), Positives = 101/104 (96%)
- Query: 1 MSTYAIITKGGKQKVEVQQAITYVEKLDVBAEAVTFNEVVLVGGRTIKVGTVPVVBGATV 60
 MSTYAIITKGGKQKVEVQQAITYVEK+D EAGAEVTFNEVVLVGG+ T VGTVPVVBGATV
 15 Sbjct: 1 MSTYAIITKGGKQKVEVQQAITYVEKIDAEAGAEVTFNEVVLVGGDKTVVGTVPVVBGATV 60
- Query: 61 VGTVEKQKQKQKVVSYKYPKKGSHRQGHROPYTKVVINAINA 104
 VGTVEKQKQKQKVV+KYPKKGSHRQGHROPYTKVVINAINA
 15 Sbjct: 61 VGTVEKQKQKQKVVTFYKYPKKGSHRQGHROPYTKVVINAINA 104

- 20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 848

A DNA sequence (GBSx0900) was identified in *S.galactiae* <SEQ ID 2565> which encodes the amino acid sequence <SEQ ID 2566>. Analysis of this protein sequence reveals the following:

- 25 Possible site: 16
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
- 30 bacterial cytoplasm --- Certainty=0.1032(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9369> which encodes amino acid sequence <SEQ ID 9370> was also identified.

- 35 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14809 GB:Z99118 excinuclease ABC (subunit C) {Bacillus subtilis}
 Identities = 221/373 (59%), Positives = 288/373 (76%)

- 40 Query: 1 MKSAAMIMHFERAAEYRDLEIAISLRTKQKRVIHQDMKDRDVFVGQVFDKGNMVCVQVFFVR 60
 M AA +EPERA E RD I I KQ++ D+ DRDVF Y DKGWMCVQVFF+R
 Sbjct: 206 MHEBAENLEPERAKELRQQLAHTESTMEKQKQMTMNDLVDRDVFAYAYDKGWMVCVQVFFIR 265
- Query: 61 NGKLIQKQVNMFPPTYNPESEDFLYIGQFYQUTKEFLPKVEVFIQDIDAKSVETIVGCKI 120
 GKLI+RDV+MFP Y E +E+PLT+IGQFY HFLPKB+ +P ID +E++ +
 45 Sbjct: 266 QGKLIERDVSMFPFLYQEADESEFLTFIGQFYQYKNNHFLPKELVPEDSIDQSMIEQLLEINV 325
- Query: 121 VKPQRGEKQLVNLAIKNARVLSQKQKFDLEKDIRKTHGAENLNLNIPKPVRIEAFD 180
 +P+G KIG+L+ LA KNA++L+KF L+E+D ++ G++ LG LNI P RI AFD
 50 Sbjct: 326 HQPKKGPKKELMLAHKQAKIALKPKSLIERDBERRSIGAVQKLGRAINITYPHIRIVAFD 385
- Query: 181 NSNIQGTSPVAAVMVFNKGKFSKKDYKFKPKITVIGPDDVYASMRVIEHRYSRVLKDLIT 240
 NSNIQGT+PV+AM+VF++GK KK+YRK+KINTV GPDDY SMREV+ RRY+RVL++ L
 55 Sbjct: 386 NSNIQGTINPVSAIVFIDGKPYKKGKRYKIKTIVTGPDDYGSREVVRKRYRVLRENL 445
- Query: 241 PPOLIVIDGQGGQVNIARDVIEHQFLAIPFIALQKNDKHQTHLELFGDPLEVVELPENS 300
 POLI+IDGG+GQ+N ARDVIEH+ GL IPFINGL K+KH+T LL GDPLEV L ENS
 60 Sbjct: 446 LPOLIIDGKGQGINAARDVIEHLEGLIPIAGLAKDKRKHSTNLIIGDPLEVAYLERNS 505
- Query: 301 BEFFLLHRIQDEVHRFAITTHQRKLSKNSPSSKLDGITGLGPKRKLQHLHGKFLSLNIQK 360

-931-

+RF+LL RIQDEVHRFAI+PHRQ+R K+*F S LD I G+G KRK+L+KHF S+ +++
 Sbjct: 506 QSEYLLRIQDEVHRFAISPHRQIRGKSAFQSVLDDIPGIGERKRLMLLKHGPGSVKMGK 565

Query: 361 AEIEDIIMGIPR 373
 A +EDI G+P+
 Sbjct: 566 ASLEIDIKHAGVPQ 578

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2567> which encodes the amino acid sequence <SEQ ID 2568>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4332 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 289/385 (75%), Positives = 334/385 (86%)

Query: 1 MKSAAMTMEFERAAEYRDLEAISLRTKQVVIHQMKDRDVFVGCMCVQVFPVR 60
 M +A+ M FERAAEYRDLI I+ +RTKQRV+ +D++DRD+FGY+VDKGWCMCVQVFPVR
 Sbjct: 206 MLAASKENAFERAAEYRDLLISGIATMRTKQVMSKDLQORDIFGYVYDKGCMCVQVFPVR 265

Query: 61 NGKLIQRDWNMFPPYFNEFEEDFLTYIGQFYQOTKHFLPKVEFTPOIDAKSVETVIGCKI 120
 GKLIQRDWN+FPYY+ EKOFILTY+GQFYQD +HF+PKVEFTP+ ID + V IV KI
 Sbjct: 266 QGKLIQRDWNLFPPYTDSEDFLTYMGQFYQKHFLPKVEFTPIEAIDEELVAIVPTKI 325

Query: 121 VKPQGEKKQLVNLAIKNARVSLQKFDLLEKDIRKTHGALENLGNLNI PKPVRIEAFD 180
 +KP+RGEKKQLV LA KNARVSLQKFDLLEKDI+ET GALENLGL LL I KPVRIEAFD
 Sbjct: 326 IKPQGEKKQLVNLAIKNARVSLQKFDLLEKDIKTSGALENLQQLRIDKPVRIEAFD 385

Query: 181 NSNIQGTSPVAAMVVFVNGKPSKDYRKPKIKTVIGPDVYASMRVHRRYSRVLKDGLT 240
 NSNIQGTSPVAAMVVFV+GKPSKDYRKPKIKTV+GPDVYASMRV+ RYRSRV K+GL
 Sbjct: 386 NSNIQGTSPVAAMVVFVVDGKPSKDYRKPKIKTVVGPDVYASMRVLPFRYSRVKKEGLQ 445

Query: 241 PPDILIVDGGQGVNIARDVIENQGLAIPAGLQKNDKHQTHLELFGDPLEVVLPFRS 300
 P+Li++DGG GQVN+A+DVIE Q GL IP+AGLQKNDKHQTH+LFG+PLEVV LPR S
 Sbjct: 446 APDLILIVDGGGVNNAKDVIEKQLGLTIPVAGLQKNDKHQTHDLLELFGNPLEVVLPFRS 505

Query: 301 EEFLLHRIQDEVHRFAITPHRLRSKNSFSSKLOGITGLGPKRKQLMKHFSLENIQK 360
 EEFLLHRIQDEVHRFA+TFHRQ+R KNSFSS LD I+GLGPKRKQLL++HFK++ I
 Sbjct: 506 EEFLLHRIQDEVHRFAITPHRVQRKNSFSSSTLDHISGLGPKRKQLLRHFKITTAIS 565

Query: 361 AEIEDIIMGIPRIVAEISLRDLSMD 385
 A E+I GIP+TV E+++ + D
 Sbjct: 566 ATSEIQALGIPKTVVAIQQQITD 590

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 849

A DNA sequence (GBSx0901) was identified in *S.agalactiae* <SEQ ID 2569> which encodes the amino acid sequence <SEQ ID 2570>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2491 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 850

A DNA sequence (GBS0902) was identified in *S.agalactiae* <SEQ ID 2571> which encodes the amino acid sequence <SEQ ID 2572>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3349 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA06651 GB:AB033763 glycerophosphoryl diester phosphodiesterase
homologue [Staphylococcus aureus]
Identities = 50/202 (24%), Positives = 96/202 (46%), Gaps = 15/202 (7%)

Query: 1 MDVIMTKDKHLVVIHDDNLIKRLSGMKNVSKITLDQVTKIPHQ---GRFA-SHIPSFTE 56
+DV +TKD +L++IHDD L+R + H+ ++++L D++ +F H+P+P+ +
Sbjct: 36 LDVAITKDEQLIIHHDDYLERITMMSGEITELANDEIKDASAGSWPGEKFKDEHLPTFDD 95

Query: 57 FMKTAQSLDQKIMIEKPY-NQNLDIYADEPIKEPKE---IRISTKHKVMSLN/TLIEK 111
+K A + + +ELK N + +K+ +E L + + + S N+ L++
Sbjct: 96 VVKIANENNNHNLNVELKIGITGPNGLALSKMVKQVEBQLNINLNQNEVLISSPNVVLVL 155

Query: 112 VEKKLPQLDTGYLIPL-----HWGTLQNH-NVDPYGIIEFSYNWIAIYLAQEYNKQLYVN 165
E+ +PQ + + W TL ++ N E+ + +E +L VW
Sbjct: 156 AEEIMPQYNRAVIPHITSFREDWRTLLDYQNAKIVNTEDAKLTKAKVMVKRAGYELNVN 215

Query: 166 TINRENLMIRYLQSPVNGIITD 187
T+N+ + V+GI TD
Sbjct: 216 TVNKFARANQLANWVGDIGITD 237
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2573> which encodes the amino acid sequence <SEQ ID 2574>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -12.26 Transmembrane 239 - 255 ( 227 - 260)
INTEGRAL Likelihood = -9.45 Transmembrane 80 - 96 ( 78 - 108)
INTEGRAL Likelihood = -9.13 Transmembrane 137 - 153 ( 131 - 160)
INTEGRAL Likelihood = -4.94 Transmembrane 278 - 294 ( 277 - 295)
INTEGRAL Likelihood = -3.56 Transmembrane 36 - 52 ( 33 - 55)
INTEGRAL Likelihood = -3.56 Transmembrane 189 - 204 ( 185 - 206)
INTEGRAL Likelihood = -3.35 Transmembrane 314 - 330 ( 310 - 331)

----- Final Results -----
      bacterial membrane --- Certainty=0.5904 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB12801 GB:Z99109 similar to glycerophosphodiester
```

-933-

phosphodiesterase [*Bacillus subtilis*]
 Identities = 67/244 (27%), Positives = 110/244 (44%), Gaps = 14/244 (5%)

5 Query: 344 VIAHRCGLVSAGVENS:FALEGRKAGSDYVRLDLILTKDNHFVSHDNRLKRLAGVNYIT 403
 +IAHRC EN++ A + A K +D +ELO+ LTED VV HD+R+ R + +
 Sbjct: 3 IIAHRCASGYAPENTIAAFDLAVKNMADMIELVQLTKDRQIVVIHDDRVDRTTNSGCFV 62

10 Query: 404 RNLTKEVEHLTSHQGH---PSGRFVSFDTPYQKAKKLAMPLELKPIGTEPGNYVDLF 460
 ++ TL+E++ L + + FG + K + LLELK ++ G ++
 Sbjct: 63 KDPTLEELQKLIVGSNYGPAPQGERIPTLEAVLKRYHKKIGLIELKGHPVSQVIGIEEVG 122

15 Query: 461 LETYHRLGSKDNKVMISLDLEVIKAKKNPSITGVYIPIQGFPG-----DEFVDP 513
 + + S+NVS ++ ++ PSI T I FG F ++
 Sbjct: 123 -QLLQFSPSINNIVQSPQPSVQRPRELYPSIFAVITRPNFGMLSRNQMKAFRSFANY 181

20 Query: 514 YVIEDFSYRSTLSSQAFWNKKIYVWINDPKRIEYLLKPKIGIITDQPALTNQLIKDL 573
 I+ + N I+ WT+N+ K + GI+TD P + +IKD
 Sbjct: 182 VNIKHRLNRLMIGSINWGLNIPAWTVNNQKTAALKQMGVDGIVTDYP---DPIIKD 238

20 Query: 574 KQDN 577
 K +N
 Sbjct: 239 IGHN 242

An alignment of the GAS and GBS proteins is shown below.

25 Identities = 90/215 (41%), Positives = 136/215 (62%)

Query: 1 MDVIMTKDKHLVVIDHDLKRLSGNKKDVSGLTLDQVTKIPHQGRFASHISPTFPMKT 60
 +D+I+TKD+ VV HD+ LKGL+G+NK + LTL +V + HCG F+ SF F +
 30 Sbjct: 375 LDLLITKDNHFVSHDNRLKRLAGVNYITRNLTLEKEVEHLTSHQGHPSGRFVSFDTPYQK 434

Query: 61 AQSGLQKIMTELKPYNQNLDTYADEFIKEFKELRLSTYKVMGLNLTLEKVEKKLQGLD 120
 A+ L+ ++IELK+ Y D F++ + L +S +KVMEL+L +IE ++KK P +
 35 Sbjct: 435 AKKIAMPLLELKPIGTEPGNYVDLPLETYHRLGSKDNKVMISLDLEVIKAKKNPSIT 494

Query: 121 TGYLIPLHWGTLQNRHFDYFGIEEPSYNWLAQRYHKLQVWVITNRDNLMIRYLQSP 180
 TGY+IP+ +G + VDFY IE+FSY +++ A NK++VWTHN + YL P
 40 Sbjct: 495 TGYIIPIQGFPGDEFVDFVIEDFSYRSTLSSQAFWNKKIYVWINDPKRIEYLLKGF 554

Query: 181 VNGIITDELNLFPKVINKKIKNSPNYQRALQLIDE 215
 + GIITD+ L + KD+K +Y+ R ++I S
 40 Sbjct: 555 IQSIITDQPALTNQLIKDLKQNSYFSRLVRIIS 589

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45 Example 851

A DNA sequence (GBSx0903) was identified in *S. agalactiae* <SEQ ID 2575> which encodes the amino acid sequence <SEQ ID 2576>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

50 INTEGRAL Likelihood = -15.02 Transmembrane 84 - 100 (76 - 112)
 INTEGRAL Likelihood = -3.50 Transmembrane 139 - 155 (139 - 157)
 INTEGRAL Likelihood = -2.23 Transmembrane 41 - 57 (39 - 59)
 INTEGRAL Likelihood = -0.96 Transmembrane 179 - 195 (179 - 195)

55 ----- Final Results -----
 bacterial membrane --- Certainty=0.7007 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

60 A related GBS nucleic acid sequence <SEQ ID 9901> which encodes amino acid sequence <SEQ ID 9902> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 2574.

A related GBS gene <SEQ ID 8671> and protein <SEQ ID 8672> were also identified. Analysis of this protein sequence reveals the following:

```

5      Lipop: Possible site: -1      Crend: 10
      McG: Discrim Score:      -3.38
      GVH: Signal Score (-7.5): -4.08
          Possible site: 53
10     >>> Seems to have no N-terminal signal sequence
      ALOM program      count: 4 value: -15.02 threshold: 0.0
      INTEGRAL      Likelihood =-15.02      Transmembrane      84 - 100 ( 76 - 112)
      INTEGRAL      Likelihood = -3.50      Transmembrane      139 - 155 ( 139 - 157)
      INTEGRAL      Likelihood = -2.23      Transmembrane      41 - 57 ( 39 - 59)
      INTEGRAL      Likelihood = -0.96      Transmembrane      179 - 195 ( 179 - 195)
15     PERIPHERAL      Likelihood = 2.01      104
      modified ALOM score: 3.50

      *** Reasoning Step: 3

20     ----- Final Results -----
          bacterial membrane --- Certainty=0.7007(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 852

A DNA sequence (GBSx0904) was identified in *S.agalactiae* <SEQ ID 2577> which encodes the amino acid sequence <SEQ ID 2578>. Analysis of this protein sequence reveals the following:

```

30     Possible site: 61
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
35     bacterial cytoplasm --- Certainty=0.4150(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 853

A DNA sequence (GBSx0905) was identified in *S.agalactiae* <SEQ ID 2579> which encodes the amino acid sequence <SEQ ID 2580>. Analysis of this protein sequence reveals the following:

```

45     Possible site: 13
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -0.32      Transmembrane      2 - 18 ( 2 - 18)

      ----- Final Results -----
50     bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```


The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 854

A DNA sequence (GBSx0906) was identified in *S.agalactiae* <SEQ ID 2581> which encodes the amino acid sequence <SEQ ID 2582>. This protein is predicted to be nad(p)h nitroreductase ydgi. Analysis of this protein sequence reveals the following:

```

10  Possible site: 38
    >>> Seems to have no N-terminal signal sequence
        INTEGRAL    Likelihood = -1.81    Transmembrane    127 - 143 ( 126 - 143)

15  ----- Final Results -----
        bacterial membrane --- Certainty=0.1723(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20  >GP:CAC09964 GB:AX033132 unnamed protein product [Bacillus subtilis]
    Identities = 62/204 (30%), Positives = 106/204 (51%), Gaps = 11/204 (5%)

    Query: 3  FLELNKKRHAVKHFNDFKVDVKTAI-KIATLAPSANNIQQMKFVVVQ--EKKSALAH 59
    F+E+ K R ++++++ K+ T I E AT APS+ N QPW+E+V+ E K LA
25  Sbjct: 7  FMEIMKRRSIRNYDFAVKISKEMTILEEATTAPSSVNAQFWRLFLVIDSPBKGKELAP 66

    Query: 60  GLPESNNQINQAYVIALFTD-----LQSRKRIARIGRRSLFDLGLYYMETLPERY 115
    L N Q+ + VIA+F D + L + K +G +P++ + L +
30  Sbjct: 67  -LASFNQITQVTSBAVIAVFADMMNADYLEETYSKAVELG--YMPQEVKDRQIAALTAHF 123

    Query: 116  ALYSEKQTGEYLSLNAGIVAMNVLVALTDQGISSNMILGFDKAITNDVLEIDK-RFREFI 174
    + E + ++ G+V+M L+L G +N I G+DK + +DK R+ F +
35  Sbjct: 124  EKLPAQVNRETILIDGGLVSMQLMLTARAHGYDTNPIGGYDKENIAETFLGKERYVFM 183

    Query: 175  LITVGYSDKVEPSYRLFVDHIE 198
    L+++G + ++ SYRLP+D I E
40  Sbjct: 184  LLSIGKAADBSYASYRLPIITIAE 207

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2583> which encodes the amino acid sequence <SEQ ID 2584>. Analysis of this protein sequence reveals the following:

```

45  Possible site: 38
    >>> Seems to have no N-terminal signal sequence
        INTEGRAL    Likelihood = -2.18    Transmembrane    127 - 143 ( 126 - 143)

50  ----- Final Results -----
        bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

55  >GP:CAC09964 GB:AX033132 unnamed protein product [Bacillus subtilis]
    Identities = 63/204 (30%), Positives = 109/204 (52%), Gaps = 11/204 (5%)

    Query: 3  FLELNKKRHAIKTNDQ-FVDYEDLRTALEIATLAPSANNIQQMKFVVVQ--BKGAELAK 59
    F+E+ K R +I+ ++ + B++ +E AT APS+ N QPW+E+V+ E K +LA

```

-936-

Sbjct: 7 FMEIMKRRSIRNYDPAVKISKEEMTRILEEATTAPSSVNAQPWRFVLVIDSPGKELA- 65
 Query: 60 GLPLA--NKVQVEQNYVVALFSDTDLALSRKIARIGV--SLPDDLGYYMETLPPRF 115
 PIA N+ QV + V+A+F+D + A +I V+ +P++ + L F
 5 Sbjct: 66 --PLASFNQITQVTTSSAVIAVFADMRNADYLEIYSEKAVELGYMPQYVKDRQIAALTARF 123
 Query: 116 AAFNEVQTGEYLAINAGIVAMNLVLSLTDQKIASNIILGPDKSTINRILDID-PRFRPEL 174
 E + I+ G+V+M L:L+ +N I G+DK E +D R+ P +
 10 Sbjct: 124 EKLPAQVNRRETLIDGSLVSMQLM+TARHGYYDTNPIGGYDKENIAETFGDKRIVFVM 183
 Query: 175 LITVGYSDKKPEPSYRLPVDVIEK 198
 L++G++ SYRLD+D +E
 Sbjct: 184 LLEIGKADBGVASYRLPIDITAE 207

15 An alignment of the GAS and GBS proteins is shown below.

Identities = 157/200 (78%), Positives = 184/200 (91%)

Query: 1 MKFLELNKKRHAVKHFNCKPVDKQVPTAIEIATLAPSANNIQPKFVVVQKKSLASG 60
 20 Sbjct: 1 MKFLELNKKRHA+K FND+PVD+D+KTAIEIATLAPSANNIQPKFVVVQKK+ LA+G 60
 Query: 61 LPESNCNQINQAVIALFTDTLQGRSRKIARIGRRSLPDDLGYYMETLPPRYALVE 120
 LP +N Q+ QACTV+ALF+DTLL RSRKIARIG +SLPDDLGYYMETLPPR+A ++E 120
 25 Sbjct: 61 LPLANKVGVQACTVVALFSDTDLALSRKIARIGVKSLLPDDLGYYMETLPPRFAAFNE 120
 Query: 121 KQTGEYLSLNAGIVAMNLVIALTDQGISSNMILGFDKAITNDVLEIDKRFRPELITVGY 180
 QTGEYL++NAGIVAMNLVLA+LTDQ I+SN+ILGFDK+ TN++L+ID RFRPE+LITVGY 180
 Sbjct: 121 VQTGEYLAINAGIVAMNLVLSLTDQKIASNIILGFDKSTINRILDIDPRFRPELITVGY 180
 30 Query: 181 SDEKVPFSYRLPVDHIEKR 200
 SDEK RFSYRLPVD +IE+R
 Sbjct: 181 SDEKFPFSYRLPVDVIEKR 200

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 855

A DNA sequence (GBSx0907) was identified in *S. agalactiae* <SEQ ID 2585> which encodes the amino acid sequence <SEQ ID 2586>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence
 40
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2895(Affirmative) < succ>
 45 bacterial membranes --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA045369 GB:U78036 dipeptidase [Lactococcus lactis]
 Identities = 312/474 (65%), Positives = 370/474 (77%), Gaps = 11/474 (2%)
 50 Query: 2 TIDFRAEVDKRRDALMDLLINLRINERDDSQADASHFPQGFQVKGLEFLEMDRQGY 61
 TIDF+ARV+KKKDALM+DL +LRI+S D ADAR+PFGQGP KAL+ FL++AERQGY
 Sbjct: 3 TIDFKAIEVKKRDLAHEDLPSLRIGDSAMDMEHADAENPFGQPRKALDAFLKIERQGY 62
 55 Query: 62 ETKNVDNYAGHFTQGGK----RELGIIFGHLLVVPAGSGNDITPYKPVKNDRNYARGSS 117
 TKN ENY GHF + G E LGI GHLLVVPAGSGND++P+RP I++ LYARG+S
 Sbjct: 63 TTKNVDNYVGHFETYGANADAELVGLIIGHLLVVPAGSGNDGNFPEIRNGNYARGAS 122
 60 Query: 118 DDKGPTACYYALKIKELGLPTSKKVRPVVGTDEESGNGMDYFFHVLGPKDPDGFSP 177
 DDKGPT+ACYALKI+KEL LP SKK+RP+VGT+EE+GW DMDYFFH LP PDGFSF
 Sbjct: 123 DDKGPTVACYALKIKELMLPLSKKIRFVGTNHEETGWADMDYFFHCLGLPLDPDGFSP 182

-937-

Query: 178 DAEFPIINGRKNITEYLHFSGEMKQAVRLHSPGSLRENMVPSATARFTSHLDQTTGL 237
DAEFPIINGRKNITEYLHSPG+G V LHSF GL ENMVPSATA + D L
Sbjct: 183 DAEFPIINGRKNITEYLHFSGKQAGVLLHSPGSLRENMVPSATAVISGAKD---LE 239

5 Query: 238 ASLADPASKH---NLKAELESVEDBOYTATVYKSAHGSTPOBVGNGATYALYLSQDFPF 294
A+L F +H NL+ +L D + T T+YKSAH+P +G+NGATYLL L+L+QDFPF
Sbjct: 240 AALEKFVARHASKNRLFDLERADKAGKATLYYKSAHGMPEKNGATYLLTFLNQDFFA 299

10 Query: 295 GPARAFDVTANIIEHDFSGEKLGVAYEDDCMGPLSNAGVFPQDETNDNTIALNFRYP 354
A AF+ V A + ED CEKLG A+ D+ M SNAGV+ FDE N + IALNFR+P
Sbjct: 300 DGAAAFIKVGAKELELDEHGEKLGTAFLVDEMENTSNAGVSPLE-NGEGKIALNFRFP 358

15 Query: 355 QGTDAKTIQTKLEKINGVEKVTLSDEHHTPHYVPMDELVSTLLAVYEKQTGLKGHEQVI 414
QG + +Q L KL+GV +V LS H HTPHYVPM D LVSTL+ VYEK TGLKG+ +I
Sbjct: 359 QGNSPERMQBELLAKLGQVEVELSKHLHTPHYVMSDPLVSTLADVYEKRTGLKGVEYTI 418

Query: 415 GGGTFGRLLERGVAYGAMFPGDENTMHQANEYMPLENI+FRSAIYARAIYELIK 468
GGTFGRLLERGVAYGAMF G+ +MHQANE P+ENI++A IYARAIYEL K
Sbjct: 419 GGGTFGRLLERGVAYGAMFPGDPMHQAENMKPVENITKAAYIYARAIYELAK 472

20

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2587> which encodes the amino acid sequence <SEQ ID 2588>. Analysis of this protein sequence reveals the following:

Possible site: 14
25 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3107(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 361/467 (77%), Positives = 403/467 (85%)

35 Query: 2 TIDFRAEVDKRRKDALMCDLNLNLRINSERDDQDAEHPFGPGFVKALEFFLEMAERDGY 61
TIDF+AEVDKRRK A++ DL++LLRINSERDD AD +HPFGPGFVKALE FL MAERDGY
Sbjct: 20 TIDFRAEVDKRRKQMLADLVLLNLRINSERDDQLDDKHFPFGFVKALEHFLAVERDGY 79

Query: 62 ETQNVENYAGHTFPQQGEELGIPGHLDDVVPAGSGMDTPYEPVIKDNRLYARGSSDDKDG 121
+T+N+DNYAG F PQGQ+E LGIPGHLDDVVPAGSGMDTPYEPVIKDR+YARGSSDDKDG
40 Sbjct: 80 KTRNIDNYAGDFEPQGEDELGIPGHLDDVVPAGSGMDTPYEPVIKDRYARGSSDDKDG 139

Query: 122 PTMACYALKIKELGLPTSKKVRFGVTDEESGNGMDYFVHVLKPKDPFGSPDAEF 181
PTMACYALKIKELGLP SKKVR+VGTDEESGNGMDYF H GL PDPFGSPDAEF
45 Sbjct: 140 PTMACYALKIKELGLPVSUKVRFGVTDEESGNGMDYFVHNLQNPDPFGSPDAEF 199

Query: 182 PIINGEKNITEYLHFSGEMKQAVRLHSPGSLRENMVPSATARFTSHLDQTTGLASLA 241
PIINGEKNITEYLHFG+G+NKA LH F GGLRENMVPSATA T+ D L A+L
Sbjct: 200 PIINGEKNITEYLHFGADNKGAFVLRHFGGLRENMVPSATAVTHDPLVLEALAE 259

50 Query: 242 DFASKGNLQAELESVEDBOYTATVYKSAHGSTPOBVGNGATYALYLSQDFGPARAF 301
F S+H +K + D + T+ GSAHGSTP+ GVNGAT LA +L+QF FEG A+ +L
Sbjct: 260 QFLSEHGKSGMKATGKIRVTIIGKSAHGSTPRAGVNGATLAKFLNQDFTFESAKOYL 319

55 Query: 302 DVTANIIEHDFSGEKLGVAYEDDCMGPLSNAGVFPQDETNDNTIALNFRYPQGTDAKT 361
V ++HEDF+ EKLG+AY DD MG LSNAGVFP FD + DNTIALNFRYP+GDTA T
Sbjct: 320 HVAGVEIHEDEFAEKLGLAYTDDMGALSNAGVFPFDQSGADNTIALNFRYPGDTDAAT 379

Query: 362 IQTKLEKINGVEKVTLSDEHHTPHYVPMDELVSTLLAVYEKQTGLKGHEQVIGGGTFGR 421
++ LEKL G+ KV+LS+HEHTPHYVPMDELV+TLLAVYEKQTGLKG+QVIGGGTFGR
60 Sbjct: 380 LKAGLEKLPGLTQVLSDEHHTPHYVPMDELVATLLAVYEKQTGLKGVEQVIGGGTFGR 439

Query: 422 LLERGVAYGAMFPGDENTMHQANEYMPLENI+FRSAIYARAIYELIK 468
LLERGVAYGAMFPGDENTMHQANEYMPLENI+RSAIYARAIYELIK
65 Sbjct: 440 LLERGVAYGAMFPGDENTMHQANEYMPLENIYRSAIYARAIYELIK 486

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 856

- 5 A DNA sequence (GBSx0908) was identified in *S. agalactiae* <SEQ ID 2589> which encodes the amino acid sequence <SEQ ID 2590>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.5598 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 The protein has homology with the following sequences in the GENPEPT database.

>GP:AC21888 GB:U32707 H. influenzae predicted coding region

HI0220.2 (Haemophilus influenzae Rd)

Identities = 123/192 (64%), Positives = 160/192 (83%), Gaps = 1/192 (0%)

20 Query: 1 MTDLKIIKAIKSDSQNYTHNGIDPLFAAPKTARINIVGQAPGLKIQBARLYWIDKSG 60
+ +L+I +I +D QN++TE GI PLF+APKTARINIVGQAPGLK +++RLIW DRSG
Sbjct: 21 LKHLDSITSSIIADPQNDFTERGIFPLFSAPKTARINIVGQAPGLFAEQSLIVNDKSG 80

25 Query: 61 DRLRQNLGVDEETFYHSGKFAVLPLDFYYPGKGGKSGDLSPRKGFAERKHPLILKEMPNVQ 120
DRLR+NLGVD + FI+SG FAVLP+DFYTPG GKSGDL PR+GFAE+WHP+IL +PN+Q
Sbjct: 81 DRLRENLGVDFYTFYHSGIFAVLPMDFYYPGKGGKSGDLFPKQFAERKHPLILKEMPNVQ 140

30 Query: 121 LTLVGGQYTKYKLGSSAHKNIETVKA YKDYLPYPLVHPSPRNQIWLKNPWFPEKDL 180
LFL+LGGY QKYL + N+T TVK Y+ +LD ++PLVHPSPRNQ+W+ KNPWF+ +
Sbjct: 141 LTLIGQAQKYLFLPEN-KDNVNTVKNYRQFLHFMFLVHPSPRNQIWLKNPWFPEQV 199

Query: 181 IVDLQKIVADIL 192

I +LQ +V I+

Sbjct: 200 IPELQILVKQII 211

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2591> which encodes the amino acid sequence <SEQ ID 2592>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3740 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45 An alignment of the GAS and GBS proteins is shown below.

Identities = 122/189 (64%), Positives = 150/189 (78%)

50 Query: 4 LKIIKAIKSDSQNYTHNGIDPLFAAPKTARINIVGQAPGLKIQBARLYWIDKSGERL 63
++ + KAI +D N +YTS GI FL+ AP+TARI IVGQAPG+ Q +L7W D+SG RL
Sbjct: 1 MDLTKAINADISANLSTYTERGIFFLYDAPQTARI IIVGQAPGTVAGQTKLYWIDRSGIRL 60

55 Query: 64 RQMLGVDEETFYHSGKFAVLPLDFYYPGKGGKSGDLSPRKGFAERKHPLILKEMPNVQLTL 123
R NLGVD +TFYHSG F ++P+DFYYPGKGGKSGDL PR+GFA RHP + MP V+LW
Sbjct: 61 RDLGLVNDUTFYHSGLPGIIFMDFYYPGKGGKSGDLFPBKGFAERKHPLILKEMPNVQ 120

Query: 124 LVGGQYTKYKLGSSAHKNIETVKA YKDYLPYPLVHPSPRNQIWLKNPWFPEKDL 183
LVG+Y Q +YLG+ A+K LIETV+ ++IDLFY PLVHPSPRNQ+WL KNPWF+DL+
Sbjct: 121 LVGRVQCPYLGKAYKYLITETVNHFFEDYLPYPLVHPSPRNQIWLKNPWFPEQLLPT 180

-939-

Query: 184 LQKIVADIL 192
 LQK V IL
 Sbjct: 161 LQKRVAIL 169

5

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 857

A DNA sequence (GBSx0909) was identified in *S.agalactiae* <SEQ ID 2593> which encodes the amino acid sequence <SEQ ID 2594>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.4178 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

20 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 858

A DNA sequence (GBSx0910) was identified in *S.agalactiae* <SEQ ID 2595> which encodes the amino acid sequence <SEQ ID 2596>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.2779 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9899> which encodes amino acid sequence <SEQ ID 9900> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD35886 GB:AB001748 conserved hypothetical protein [Thermotoga maritima]
 Identities = 36/124 (29%), Positives = 58/124 (46%), Gaps = 3/124 (2%)

40 Query: 19 VPTKELLADYFNNRMEFAIGRVEAHVLAHFDYGPRIKILDOVEDLKEPFTQLKRIFIKMLSK 78
 +P EL DY R F + RV+ H LAH DY R D K +++I + +
 Sbjct: 96 LPPDELARDYLERTLFVMERVKPHTLAHLDPARYAKAD---FRANRDLEKILVELVGN 154

45 Query: 79 GLAFELNTKSLYLTYGNEKIYKYALEILKQLGCKQYSIGSDGHIPEHPCYEFDRLLQCLRD 138
 A E+NT L+ +G + +E+ LG + +IGSD H +H + + LK
 Sbjct: 155 EKALEINTAGLFGKNGKPNPDYIVEMTYDLGGRVVTIGSDAHISCHIGRGIEEVMRELK 214

Query: 139 YQID 142
 + +
 50 Sbjct: 215 FNP 218

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 859

A DNA sequence (GBSx0911) was identified in *Sagalactiae* <SEQ ID 2597> which encodes the amino acid sequence <SEQ ID 2598>. This protein is predicted to be alkaline amylopolulanase (pulA). Analysis of this protein sequence reveals the following:

```

Possible site: 41
>>> Seems to have no N-terminal signal sequence
10  INTEGRAL    Likelihood = -10.08    Transmembrane 1225 -1241 (1222 -1247)
    INTEGRAL    Likelihood = -2.44    Transmembrane 19 - 35 ( 18 - 36)
    INTEGRAL    Likelihood = -0.11    Transmembrane 1146 -1162 (1146 -1162)

----- Final Results -----
15  bacterial membrane --- Certainty=0.5034 (Affirmative) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20  >GP:AAG33958 GB:AF217414 pullulanase [Streptococcus pneumoniae]
    Identities = 641/1311 (48%), Positives = 854/1311 (64%), Gaps = 88/1311 (6%)

Query: 1 MRRKDLFGDKQTQYTIKRLSVGVASVATGVCIFLHSPQVPABEVSASPANTAIAESNINQ 60
      M++ +K+ Y+IR L G SV G + L A+A I+
25  Sbjct: 1 MKRTPSHTEKKMVIYSIRSLKMGTVGSVLIGASLVL-----LMMATPTISS 44

Query: 61 VDNQQTSLNKDDINSETVVTFSIMPEPTKQLVSDSTDTQKGVTRPKDKATSLLEENKQ-P 119
      ++ +TN + N N+ T+ P + DT + ++ P A + LE+ + P
Sbjct: 45 DESTPTTN--EPNNRNTITLAQP--LITD---AAGSGNSEDSSPGNANASLEKTEEP 97

30  Query: 120 VSDIKYLLDKVAPSTLQNTCPDKTSQAIGAPSPITLKVANQAPRIENGYFRHLHLKEPQGH 179
      ++ T A Q D+S+ + LP + IE+ YFR+H+K+LP +
Sbjct: 98 ATEETTPAASPADPAQCQDQRSEPTTSTSPVTTETKASEPIEDNYFRIRHVKKLEPEENK 157

35  Query: 180 VESTGLWINGDVPDPSNPNPAGIPIWTDKAKDDYGYTDFPKLSEKQKQISFLINNAGT 239
      ++ GLW W DV++PS NWPNGA+ DAKKDDYGYD+ KL +Q K +ISFLINN AG
Sbjct: 158 -DAQGLWTDNDVEKPSNPNPAGALSFKCAKKDDYGYLDVCLKGKQAKKISFLINNAGT 216

Query: 240 NLSDHHLPLLRPEMNQVWIDKQYGTHTYQPLKEGYVRINYLSSSNYDHLAWFLKDV 299
      NL+GD + L P+MN+ W+D+ Y +Y+P G VR+NY + NYD S W +D
40  Sbjct: 217 NLTDGRSEVKL/PIONEAHLQDQYKVSFEPQAGTVR/VNYRTDGNYDKSLWYGVGVK 276

Query: 300 TPSTT-WPDGSNPFVNGLYGRYIDVSLKTNKAIGFLILDSKTDGAVKVPNDYVFRDL 358
      PS+ WPDG++F G YGYTD+ L A+E GFL+LDESK GD VK++ +Y F DL
45  Sbjct: 277 NPSSAQWPDGDTPTATGKYGRYIDPLNEAAREFGFLILDSKQGDVVKIRKNTYKFTDL 336

Query: 359 ANHNQIFVKDKDPKVINNNPYIDQVLQKDAQQIDLTSLQASPTFLDGVKTEILKELKVT 418
      NH+QLF+RD D +Y NPYT+ +++ AQ + +SI++SF+TL G K +ILK +T
Sbjct: 337 KNSQIFLKKDDDESIYTNPYTHDTRTKAQHVGTSSISSFSFLVGAKKEDILKHNIT 396

50  Query: 419 DRKNQATQISDITLDTSKSLIILKODPNPKQGHFNI SYGNKNVMTQSWEPFDQLYAYSG 478
      + N + I+D+ +D + + GDF+ + + +SYN + T+ SW ED+ Y+Y G
Sbjct: 397 NHELGNKVTITDVAIDEAGKKVYSGDPSCTKHPTYTSVNSDQFTTKTSMRLKDETYGYDG 456

55  Query: 479 NLGAVLNQDGSKEVSLWSPGSDAVTMIYDKDNQNRVVAITFLNNQNGVWQITLTD -- 536
      LGA L ++G +V+ +LWSPSAD V++++YDK++ ++VV T L K +G W+ LD+
Sbjct: 457 KLGAFLKBEKGQVDLTLWSPSADKVSVVYVDKNDPDKVGTVALEKGERGTWQYLDSTN 516

Query: 537 KLGIKNTYGYYYLYEIKRGKDKVKILDFYAKSLAENDSNT--VNDDIKTAKAAFPVPSQL 594
      KLGI ++TGYYY Y+I+R V LDFYAKSLA W+S+ ++D K AKAAPV+P++I,
60  Sbjct: 517 KLGITDFYGYYYQYIBRQKTVLALDFYAKSLA WNSDDAKIDARKVAQAAPVDPAKL 576

Query: 595 GPQNLSPAKIANFKGRQDAVIYEAHRVDFTSDRLSGLKNQFGTTPAAPSRLDYLLQKLG 654

```

GPQ+L++ KI NPK R+DAVIEAHVRDFTSD ++ L PGT AF EKLDYL+ LG
 Sbjct: 577 GPQDLTYGKIHPKTRDAVIEAHVRDFTSDPAIAKDLTKPCTPFAPIEKLDYLLKLG 636
 Query: 655 VTHIQQLFVLISYFYVNMEDKSRSTA-YTSSDANNVNGYDQPSYFALSGMYSEKPDPSAR 713
 5 VTHIQQLFVLISY++VME+ + Y SS++NVMNGYDQPY+L+GMYSPK+P R
 Sbjct: 637 VTHIQQLFVLISYFYVNEIKNHEHLSYASSNNYNGYDQPNFYSLTQMYSSDPNPKER 696
 Query: 714 IAEKQLLTHDTHKRGMGVILDVVYNHTATLYLFDIEFNVYHFMEDGSPRESFGGGRG 773
 10 IAE K LI++THKRGMG ILDVVYNHTAK +FED-EPNYHFM+ DG+PR SFGGGRG
 Sbjct: 697 IAEFNKLINEHKRGMAILLDVVYNHTAKVDIFEDLKFNYHFMADRGTPKTSFGGGRG 756
 Query: 774 TTHMSPRVLVDISIKYLTSEFKVDGPRFDMMDHDAADAAIEAYKBAKAINPMNIMIGEGW 833
 TTH M++R+LVDSIKYL +KVDGPRFDMMDHDAADAAIEAYKBAKAINPMNIMIGEGW
 Sbjct: 757 TTHMTKRLVLVDISIKYLVDTYKVDGPRFDMMDHDAADAAIEAYKBAKAINPMNIMIGEGW 816
 Query: 834 RTFGSDQQQPVKPADQDMWKSTDTGVFSDDIRNSLKSQFPMGTPAFITGGPSLQGIT 693
 RT+ GD+ P K AQDMMK TDTV VSSDIRN+LKSQ+PMG PAFITGG + + IF
 Sbjct: 817 RTYAGDENMPTKAAQDMWKHTDTVAVFSSDIRNLSKSGTFMGCQAFITGGKRDVNTIF 876
 Query: 894 KKHKAQPNFADSPGDVVCYIAAHNLTLDHVIASINKNDPKVBE--EIHRRRLGNV 951
 20 KH+ AQF NFEADSPGDVVCYIAAHNLTLDHVIASINKNDPKVBE--EIHRRRLGNV
 Sbjct: 877 KHLIAQPTNFADSPGDVVCYIAAHNLTLDHVIASINKNDPKVBE--EIHRRRLGNV 936
 Query: 952 MILTSQGTAFIHSGQCYGRTKRLNPDYMTKVSDDKLPNKAILLEAVK---EYPTFIHD 1007
 25 M+LZ+GTF FHSQCYGRTK+ NP Y T V++DK+PK+ L+ +YPTFIHD
 Sbjct: 937 MVLTAQSTPFIHSGQCYGRTKQFRNPAYRTFVABDKVINKSHLLRDKDGNPDYPTFIHD 996
 Query: 1008 SYDSDDAINHFDMAAATDNNKHPISTKTQATYAGLITLRESTDAPRKLKAEIDREVSIL 1067
 30 SYDSDDA-N FDM ATD +P+K++ Y GLI LR+STDA PR S +I V LI
 Sbjct: 997 SYDSDDAVNKPDWTKATDQKAYPENKSRDYMGLIALRQSTDAFRLESQIDKRVHLI 1056
 Query: 1068 TEVQCGDIKEKDLVIAQTIDSKEDIYAVFNADSKARNVLGKRYKHLKQGVIVADQ 1127
 T GQ +++D+VI YQ GDIYAVFNAD KAR LG +HL +V+ D+Q
 Sbjct: 1057 TVPQNGVEKEDVVGICITAPNGDIYAVFNADKAREFNLTGTAFAHLNRAEVLADENQ 1116
 Query: 1128 AGIKPISTPRGVHPEKDSLLIDPLTAIVIKGVKAPS-----PKELQAD 1172
 35 AG I+ P+G+ + + L++ LTA V++V + S P+ +A
 Sbjct: 1117 AGSVGIANPKGLEWTEKGLKLNALTATVLRVQNGTSHETAEEKPDSTSPKPEKQREAS 1176
 Query: 1173 YPKTQ-----SPKESKTEVKVNRANKT-----SITPVVSKKADS 1207
 40 +P Q + +K+ N+ + T S+ V K++
 Sbjct: 1177 HPAHQDPAPEARPDSTKPKAKVADAENKPSQATADSQAQCPAQEAQSSVKEAVERKESVE 1236
 Query: 1208 YLTNE-----ANLPKTDKSSKILSVVGIILASLLAVGLSLKRN 1249
 45 + E A LP TG K+ L GIS+LA LL LG LK +
 Sbjct: 1237 NSSKENISATPDAQELPWTGINKENKLLPAGISLLA-LLGL-GPLKNNK 1285

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2599> which encodes the amino acid sequence <SEQ ID 2600>. Analysis of this protein sequence reveals the following:

50 Possible site: 35
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -10.83 Transmembrane 1153 -1169 (1148 -1171)
 INTEGRAL Likelihood = -1.97 Transmembrane 29 - 45 (28 - 46)
 55 ----- Final Results -----
 bacterial membrane --- Certainty=0.5331 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

60 A related sequence was also identified in GAS <SEQ ID 9125> which encodes the amino acid sequence <SEQ ID 9126>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 26
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty- 0.533(Affirmative) < succ>
bacterial outside --- Certainty- 0.000(Not Clear) < succ>
bacterial cytoplasm --- Certainty- 0.000(Not Clear) < succ>

5

LPXTG motif: 1133-1137

An alignment of the GAS and GBS proteins is shown below.

Identities = 715/1097 (65%), Positives = 872/1097 (79%), Gaps = 21/1097 (1%)

Query: 156 ANQAPRIENGYFRHLKSLPQGHFVSTGLWINGDWDQPSNWPNALPTMDAKDDYGY 215
AN A E+ + R+H K LP G + S GLW+NSDWDQPS +WPNAGI NT AKDDYGY
Sbjct: 95 ANPASIASH-HLRMHFKTLPAESGLSGLWVWGDVDQPSKDWPNAGIITTKAKDDYGY 153

Query: 216 YVDFKLSEKQRKQISFLINNKAGTNLSGDHII PLLRPEMNQVWIDHGYGTHTYPLKSGY 275
Y+D L+ K R+Q+S-LINNKAG NLS D HI LL P+M+VWIDE Y H Y+PLK+GY
Sbjct: 154 YLDFPLAAKHRCQVSTYLLINNKAGNLSGDHII SLLTPMNEVWIDENTHATATPLKNGY 213

Query: 276 VRNTYLSSSSYDHLASAMFKVDATPSTTWPDGNSFNQCLKGRVIDSLATNAKELGY 335
+RINT + S +YD+L+ M FKDV TP+T WP+G + ++G YG Y+DV LK A EIGFL
Sbjct: 214 LRINTHNGSGHYDNLAWTFYKDVKTPTTWPNGLDLSHGHYGATVDVFLKEGANEIGFL 273

Query: 336 ILDESKTGDVAVKQPNDFVFRDLANNKQIFVKDKDPKVNNPFTYIDQVLKDAQIDILTS 395
ILD+SKTGDA+KVQP DY+P++L NH Q+PVKD DPKVNNPFTYIDQV LK A+Q
Sbjct: 274 ILDESKTGDALKVQPKDYLFKELDNHTQVFKVDTPKVNNPFTYIDQVSLKGAECITPNE 333

Query: 396 IQASFTTLGDVDETEILKGLKVIDKNQNAIQISDTILTSKLLIKGDPMKPGQHFNGL 455
I+A FTTLDG+D+ + + +K+TDK + I + +TLD KS+ +KGF + + + +
Sbjct: 334 IKAFTTLGLDVEDAVKQIKITIKTDEAGKTVAIDELTLDKDSVMFLKGDFAQAGAVTTP 393

Query: 456 YNGRVMTRQSEWPKDQLVAYBNCGLAVLMDQSGKVASLMSPSADSVMIITVDKDNQR 515
+ + + RGSW KD+LHAY G LGA L +DGS V+ +LMSPSAD+V +++YDK +Q R
Sbjct: 394 PGEVSQVARGSWQLDKLKYAYDGLGATLADKDS-VDLALMSPSADTVKVVYTDKQDTR 452

Query: 516 VVATPTLMKNNKGNVQITL--DTKLRIHTITGYYYLYEIKRGDKVKILDPYAKSLAEN 573
VV L K+ +KSW+ L D+ GI +YTGYYLYEY RG++IV +LDPYAKSLA N+
Sbjct: 453 VVGQADLTSDKGVWRAHLTSDSVKGISDITGYYYLYEITRQGERWMLDPYAKSLAEN 512

Query: 574 SNTVNDIDIKTAAAFVNPSPGLPQNLSFAKIANFKRQDAVIYEAHVDPFTSDRLDGL 633
T DDIKTAAAF++PS+LGP L FAKI NFK R+DA+TYEAHVDPFTSD++L+GKL
Sbjct: 513 DATATDDIKTAAAFIDPSKLGPTGLDPAKINPKREDATYEAHVDPFTSDKALEGL 572

Query: 634 KNQPGTFAAFSEKLDYLLQKLGVTIQLLPVLGYFVYVNMKDSRTATSTDNNTYNGWYD 693
+ FGTP+AF E+LDYL+ LGVTH+QLLPVLGYF NE+DKSRSTATSTDNNTYNGWYD
Sbjct: 573 THPPGTFAFPEQLDYLLQKLGVTIQLLPVLGYFVYVNAELDKSRSTATSTDNNTYNGWYD 632

Query: 694 QSYFALSGMYSEKPKDPSARIABLQHLIHIIHKRGMGVILDVVYNHIAKTYLFDLEPNY 753
Q YFALSGMY S PD+ RIABL L+ ++IHKRGMGVI DVVYNHIA+TYLFD+EPNY
Sbjct: 633 QSYFALSGMYSANPIDPALRIABLQHLVNEIHKRGMGVIPDVVYNHIAKTYLFDLEPNY 692

Query: 754 YHFMNEDGSPRESFQGGRLGTHANSRRVLVDSI KYLTSKFKVDGFRFPMKGDHDAAE 813
YHFMN DG+ RESFQGGRLGTHANSRR+LVDSI YLT EFKVDGFRFPMKGDHDAAE
Sbjct: 693 YHFMNADTARESFPQGGRLGTHANSRRILVDSITTYLTFKVDGFRFPMKGDHDAAE 752

Query: 814 LAVKENKAINPIMIIGENKTYTQGDQCPVKPADQWMMKSTDTQVPSDDIRNLSKGF 873
A+K AAKAINP IMIIGENKTYTQGD+G+ ADQWMMK+T+TVGVFDDIRN+LKSGF
Sbjct: 753 QAFKAAKAINPIMIIGENKTYTQGDGKKISADQWMMKATNTQVPSDDIRN+LKSGF 812

Query: 874 PNEGTPAFITGGPQLGHI FKNIKAQGNFEADSPGVQVYIAAHNKLTHIVIAKSINK 933
PNEGTPAFITGG +L+G+PK IKAQGNFEAD+PGVQVYIAAHNKLTHIVIAKSINK
Sbjct: 813 PNEGTPAFITGGAKNLESLPKTIKAQGNFEADAPGVQVYIAAHNKLTHIVIAKSINK 872

Query: 934 DPKVAEEIHRRLRGLNMVILSQTAFIHSQGYKGRTRKLLNPDMYTKVSDKLPNKAT 993
DPKVAEEIHR+R+RLGN MIIT+QGTAFIHSQGYKGRTRKLLNPDY TK SDK+PNKAT
Sbjct: 873 DPKVAEEIHRRLRGLNMVILTAQTAFIHSQGYKGRTRKLLNPDYTKVSDKLPNKAT 932

Query: 994 LIEAVKEYPYFIHDSYDSSDAINHFDWAAATNNKEPISTKTQATYAGLTLRSTDAFR 1053

		Li+AV +YPVFHDSYSDSDA+NHFDWA ATD+ HPIS +T+ATT GLI IRRSDAF	
	5	Sbjct: 933 LLDVAQYTPFVSLHSDYSDSDA+NHFDWA KATDS LAHPISQTKATQGLIAIRSDAFT	992
		Query: 1054 KLSKAKPRDESVSLHSDYSGQSDIKKKDLVATYQT+IDSKGDIYAVFNADSKARNVILGEKY	1113
		K+KAE+DR+V+LTT+ GQ I++DL+ YQT+ S GD YAVFNAD+K R V+L+P	
		Sbjct: 993 KATKAEVDRI+LTT+GQ I++DL+YQT+ S GD YAVFNAD+K R V+L+P	1052
		Query: 1114 KHLLEGQVIVDADQAGIKPSTPRGWHFPEKESLLDPLTAIVIKV-GKVAPSKKEELQAD	1172
		+LL QV+YDA+QAG+ I+ P+GVP F K+ L+ L+ LTA+V+K K A+ + + Q D	
10		Sbjct: 1053 RYLLGAQVYDAGQAGVTAIAKPKGVPTKBLTBSLTAIVIKVSKSTANPSSQKSTD	1112
		Query: 1173 YPKTQSKESKTVZKVNRIANKTSITPVSKKADSYLITEANLPKTDKSSKSLVVGIS	1232
		+T++ SK ++K K + T LPKTG+ SSK L GI+	
		Sbjct: 1113 NHQKTPDGSKDLKSLMTRPKRAKT-----NQKLPKTGESSKSLLAAGIA	1159
15		Query: 1233 ILASLLAVGLSLKRN 1249	
		+ LL + L +KR +	
		Sbjct: 1160 L---LLLAISLLMKRK 1173	

20 A related GBS gene <SEQ ID 8673> and protein <SEQ ID 8674> were also identified. Analysis of this protein sequence reveals the following:

```

25 Lipop: Possible site: -1   Crend: 9
    M6G: Discrim Score:     -0.88
    GVH: Signal Score (-7.5): 4.13
        Possible site: 41
    >>> Seems to have no N-terminal signal sequence
    ALCM program   count: 3 value: -10.08 threshold: 0.0
        INTEGRAL Likelihood =-10.08   Transmembrane 1225 -1241 (1222 -1247)
    30 INTEGRAL Likelihood = -2.44     Transmembrane 19 - 35 ( 18 - 36)
        INTEGRAL Likelihood = -0.11   Transmembrane 1146 -1162 (1146 -1162)
        PERIPHERAL Likelihood = 2.44   653
    modified ALCM score: 2.52

    *** Reasoning Step: 3

    ----- Final Results -----
        bacterial membrane --- Certainty=0.5034(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    40 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

    LEXYNG motif: 1081-1085

```

The protein has homology with the following sequences in the databases:

```

45  GRP00953|1111 - 3768 of 4356)
   RGD|165156|TM1845(18 - 840 of 843) pullulanase {Thermotoga maritima}[sp|033840|PULA_THEMA
   PULLULANASE PRECURSOR (EC 3.2.1.41) (ALPHA
   -DEXTRIN      ENDO-1,6-ALPHA-          GLUCOSIDASE)          (PULLULAN      6-
   GLUCANGLUCOYLASE).GP|2815006[emb|CA04522.1|][AJ001087 pullulanase {Thermotoga mari
   tina}[GP|4982428[gb|AAD36907.1|AR001821_7|AR001821      pullulanase      {Thermotoga
50  maritima}[PIR|H72204|H72204 pullulanase - Thermotoga mariti
   ma (strain MSB8)
   %Watch = 8.4
   %Identity = 30.6 %Similarity = 52.8
   Matches = 210 Mismatches = 298 Conservative Sub.s = 152

55  1032      1062      1092      1122      1152      1182      1212      1242
   NKAGNLSGDHPLLRPMNQVWIDREKYTHYTHYPLKGGVVRINYLSSSENVDHLSAWLFDVATPSTTTPDGSNEVNG
   |      |      |      |      |      |      |      |      |      |      |      |      |
   MKTKIMLVLLVLLLSALIPSETTVIIVHYHRYDGYDGNLWLP--VEPVSQBGKAYQFTGE

60  10      20      30      40      50
   1272      1302      1329      1359      1668      1698
   GLYGRYIDVSLKTNKEIGFLI-LDSKTKGDAVKVQPNIDYFRDLA-----EQCGHFNISYGNNNVMTQSGMEFKDL--
   |      |      |      |      |      |      |      |      |      |      |      |      |
   DDGPGKAVVGLKMLTKVGLIVRLN-----WQKGVQAKDR
65

```

	70	80		90
	1746	1776	1806	1836
5	-----XAYSGNLGAVLNQDGSKVRASLWSPADSVMIIYDKDN			
	FIEIKDGRAEVWILQGV-----ELITEGYKPARVIMMEILDVYYDRELGA VYSFE--KTIIFVWSPFSKWKVLLFKNGE			
	110	210	220	230
	240	250		
10	1866	1896	1926	1956
	1986	2016	2046	2076
	QNRVATTPLMKNKGVWOTILDITLGIKNYTYGYLYEIKRGRKVKILDPYAKSLAWDSMTVNDIDITAKAFAFVNP			
	DTEPYQVWMEYKNGVWZAVVBDL-----DGVFLYQLEMYGKIRITVDPYSKAVIA-----NSKKSAVVNLA			
	270	280	290	300
	310	320		
15	2106	2136	2166	2196
	2226	2253	2283	
	QLGPNLSPAKIANFQRQDAVIYEAHVRDPTSDRLDGLKNQFGTFAPASEK-----LDYLQKLGVTIRIQL			
	RTNPEGWENDRGPKIIGYEDAIIVEINIADITG--LENSGVIONK-GLYGLTERTNKGPGGVITGLSHLVELGVTIRVNL			
	330	340	350	360
	370	380	390	
20	2313	2343	2373	2403
	2433	2463	2493	
	PVLGYTFYNMIGBSRSTAYTSDNTNYMGYDQGYFALSGMYSEKPKDPSARIELKQLINDHKRGMGVILDVVYNT-			
	PFDFPTDDELQK-----DFEKYNYMGYDPTLPMVPEGRYSTDPKPHTRIREVKEVMVLAHKGIVMDVVFPHY			
	410	420	430	440
	450	460	470	
25	2544	2574	2601	2631
	2661	2691	2721	2751
	---AKTYLPEIDPNHYHFWNDGSP-RESFGGRGLGTHAMSRRLVDSIKYLSEFKVDGFRFMDGHDHAIELAYK			
	GIGELSAFDQTVPPYFIRIDITGAYLNESSGCVIASERPMMKRFIVDTVTYVWKEYHIDGFPOQMLIDKKTMLEVER			
	480	490	500	510
	520	530	540	550
30	2781	2811	2841	2871
	2901	2931	2979	
	EAKNINFMIMIGBSWRTFQDQGGFVKPADQWKKSTDTGVGSDDIRNSLKGSGFFNBSGTPAFITGG-----PQSLQSI			
	ALHKIDPTIILYGEFW-----GGWGAIPRPKSD--VAGTHVAFNDEFRDAIRGSVFNPSVKGPFMGVYKSTKIKRGIV			
	560	570	580	590
	600	610	620	
35	3030	3060	3084	3114
	3144	3174	3204	
	KNIAQPG-----NFEDSGPDVVQYLAANDNLTHD--VIAKSINKDPKVAEEIHRRLRLGNVMLTSQTAIFSHSQEY			
	GSINDYGLKIVSFALD-PEETINYAACHDNHILMDKNYLAADKADKKKEWTEELAKNAQKLAGAILLTSQGVFFLHGQDF			
	640	650	660	670
	680	690	700	
40	3234	3264	3294	3324
	3354	3384	3414	3444
	GRTRKLLNPDMYNTSVSDDKLPNKATLEAVKEYPPIHDSYDSSDAINHFDMAAATDNNHPISTKTQYTAGLITRRS			
	CRITN-----FNDNSYNAPISINGFDY-----ERKLPIDVFNHYKGLIKRKE			
	710	720	730	740
50	3474	3504	3534	3564
	3594	3624	3654	
	TDAFRKLSKAEI-----DREVSLLITEVGGDIKEKDLVIAYQTIIDSKGDIYAVFVNADSKARNVLLGEKYKHLK			
	HPAFRLKNAEIKKHLFPLPGGRRIIVAPMLKHAGGDPMKQIVVIYN-----GNLEKITYK-LPE			
	760	770	780	790
	800			
55	3678	3708	3738	3768
	3798	3828	3858	3888
	GQ--VIVDADQAGKPKISTPRGVIFEKDGLLIDPLTAIVIKGVKVAPEPKSELQADYPKTQSPKESKTVEKNRIIRIAKTS			
	GKNWVVNSQKAGTEVETVEG-----TIELOPI SAYVLYRE			
	820	830	840	
60				

SEQ ID 2598 (GBS5) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 3 (lane 7; MW 134kDa).

The His-fusion protein was purified as shown in Figure 190, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 860

A DNA sequence (GBSx0912) was identified in *S. agalactiae* <SEQ ID 2601> which encodes the amino acid sequence <SEQ ID 2602>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -10.72 Transmembrane 231 - 247 ( 228 - 251)
INTEGRAL Likelihood = -8.39 Transmembrane 50 - 66 ( 44 - 68)
INTEGRAL Likelihood = -6.74 Transmembrane 23 - 39 ( 20 - 41)
INTEGRAL Likelihood = -5.84 Transmembrane 173 - 189 ( 168 - 196)
INTEGRAL Likelihood = -4.41 Transmembrane 299 - 315 ( 297 - 318)
INTEGRAL Likelihood = -4.14 Transmembrane 115 - 131 ( 114 - 133)
INTEGRAL Likelihood = -3.35 Transmembrane 80 - 96 ( 79 - 97)
INTEGRAL Likelihood = -0.48 Transmembrane 97 - 113 ( 97 - 113)

----- Final Results -----
bacterial membrane --- Certainty=0.5288 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8675> which encodes amino acid sequence <SEQ ID 8676> was also identified. Analysis of this protein sequence reveals the following:

```

SRCFLG: 0
MoG: Length of UR: 19
Peak Value of UR: 3.08
Net Charge of CR: 1
MoG: Discrim Score: 9.76
GvH: Signal Score (-7.5): -4.57
Possible site: 22
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program count: 7 value: -10.72 threshold: 0.0
INTEGRAL Likelihood = -10.72 Transmembrane 217 - 233 ( 214 - 237)
INTEGRAL Likelihood = -8.39 Transmembrane 36 - 52 ( 30 - 54)
INTEGRAL Likelihood = -6.74 Transmembrane 9 - 25 ( 6 - 27)
INTEGRAL Likelihood = -5.84 Transmembrane 159 - 175 ( 154 - 182)
INTEGRAL Likelihood = -4.14 Transmembrane 101 - 117 ( 100 - 119)
INTEGRAL Likelihood = -3.35 Transmembrane 66 - 82 ( 65 - 83)
INTEGRAL Likelihood = -0.48 Transmembrane 83 - 99 ( 83 - 99)
PERIPHERAL Likelihood = 0.26 136
modified ALOM score: 2.64
icml HYPID: 7 CFP: 0.529

*** Reasoning Step: 3

----- Final Results -----
bacterial membrane --- Certainty=0.5288 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB08178 GB:AB036768 exfoliative toxin A [Staphylococcus hyicus]
Identities = 134/298 (44%), Positives = 197/298 (65%)

Query: 22 PLVAGLVLGLLGLNLLEGGVGYTYRYCLGLNALVFWLPLKGLIKKKKSKRKELGNPLI 81
          PLV +GLVGLGL LGNLL+ + G+AA+ W+ L+ + N + +L+PL+
Sbjct: 7 PLVSGGLVLGLLGLNLKDVSLGLNALCGLAILWHLILLYGMFNFNIVKNGQLNSPLV 66

Query: 82 ASVPTTFNAGMILSTYILLFRSLGIWVAVLKGVWMLSPALILHMAIFSMKYLRFHSM 141

```

-946-

+SVPTTFPM+G +TY+ F S ++ L +W L I ++ HM IFS KYL+ FS+
 5 Sbjct: 67 SSVPTTFPM+G+LGTYYTATPFSHISPIHILITPMLLCLIGLIMHIIIFSXYKIGLPSL 126
 Query: 142 ANLPFGMSVLYVGIVASLTAPISGQPTIGKIVFWMGFIATLWLPFLFIKAYKIGLPSA 201
 N++PSM+VLY+GI +A LTAP+SG F IGR+ YGF+AT ++LP +F + L ++
 Sbjct: 127 ENVYPSWVLYIGIALAGLTAPVSGYFFIGKLVYGFVATCIVLPLVFKRKYLPQTFS 186
 Query: 202 VKNPTTTCAPMSLITAGVYNSFVSPNRGLLLIIVMAQFLYFFILFQVSKLLIGDPTG 261
 +KFN +TICAP SL+ A YV +F + +++L +++Q YF+I+Q+KILL F+P
 10 Sbjct: 187 IKPNTSTICAPSLVAAYVLAFFHMDVVFILILSQQVFFYITVQPKLKRFPSPV 246
 Query: 262 FSAPTFLVISATSLKLSLQHLSPVDIOGLVHIFEGITTLIVMVMVRYIFFPRT 319
 FSAPTFLVISAT+LK S+ L P GL+ FS T+IV V YI ++
 Sbjct: 247 FSAPTFLVISATLQSMFLIPPEINWGLMFETVLTATVIVFVFGYIHLKPV 304

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2603> which encodes the amino acid sequence <SEQ ID 2604>. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence
 20 INTEGRAL Likelihood = -9.82 Transmembrane 169 - 185 (163 - 189)
 INTEGRAL Likelihood = -8.49 Transmembrane 50 - 66 (38 - 69)
 INTEGRAL Likelihood = -7.86 Transmembrane 228 - 244 (224 - 247)
 INTEGRAL Likelihood = -5.15 Transmembrane 288 - 304 (284 - 306)
 INTEGRAL Likelihood = -3.29 Transmembrane 108 - 124 (107 - 126)
 25 INTEGRAL Likelihood = -3.29 Transmembrane 140 - 156 (140 - 161)
 INTEGRAL Likelihood = -1.33 Transmembrane 84 - 100 (84 - 100)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4927 (Affirmative) < succ>
 30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 138/305 (45%), Positives = 200/305 (65%), Gaps = 5/305 (1%)
 35 Query: 12 RYMKGWEKPLVMAQLVLGLLGNLSEGYTVYRCLGVALVWFIKIGLIRKKE 71
 R +WK+ + PRLVM+GL LG L+ GNLL Y + Y L AL + L+ G++N+
 Sbjct: 12 RTMKGHLKTPPLVMSGLALGTLSPGNLLATVYSINVLQIALLPIYGILLVGMVRNLD 71
 40 Query: 72 SRKELSNPLIASVPTTFPMAGMLSTYILLFRLGIVFVAVLSKGVWMLSPIALIIMHAI 131
 ++ +L PLIASVF TFM GM+LG+ L G W+ L+ WML P+ ++ A+
 Sbjct: 72 TKQQLRQPLIASVPTTFPMGMLSSFLKVTG-GCNLGFLT---WMLPLGNVLIIAYY 127
 45 Query: 132 SMKYLHPSMANLPFSMSVLYVGIVASLTAPISGQPTIGKIVFWMGFIATLWLPFLPT 191
 ++++ PS N+PSMSVLY+VGI +A LTAP S QF +G+++FW + T V+LFF+
 Sbjct: 128 QYRFVFSFMDNVPPSMGVLVFGIAMAALTAASRQPLLQGVIFVWCLLTAIVLFPNAK 187
 Query: 192 KAYKIGLPSAVKFNPTTTCAPMSLITAGVYNSFVSPNRGLLLIIVMAQFLYFFILFQV 251
 K Y IGL AV FNI+T CAP+SL++A Y+ +F P G+++ L+V F++ Q+P
 50 Sbjct: 188 KTYGIGLQAVHNPSTTCAPSLLSASYLATPRPQGVNFIPLVSSQLLYAVVWQLP 247
 Query: 252 KLLIGDPTFGSAPTFLVISATSLKLSLQHLSP-VDIOGLVHIFEGITTLIVMVMVR 310
 +LL P PGFSAPTFF VISATSLK++ L + Q L+ E+ T+V + V
 Sbjct: 248 RLNRFPNPGFSAPTFFVISATSLKMTLSFLGWSGLQVLLGVEVLTALVTVVGA 307
 55 Query: 311 YIFFL 315
 Y+ FL
 Sbjct: 308 YLRF 312

60 Based on this analysis, it was predicted that these proteins could be useful antigens for vaccines or diagnostics.

Example 861

A DNA sequence (GBSx0913) was identified in *S.agalactiae* <SEQ ID 2605> which encodes the amino acid sequence <SEQ ID 2606>. Analysis of this protein sequence reveals the following:

```

Possible site: 28
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2607> which encodes the amino acid sequence <SEQ ID 2608>. Analysis of this protein sequence reveals the following:

```

Possible site: 28
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 45/57 (78%), Positives = 53/57 (92%)

Query: 1  MVKKFPAFKGIATGVVATAATLAGAFAIKKTIIPREEKIAFIEENRKKAAARKRVS 57
      F  KG+ATGV+ATAAT+AGA  FA+KKTII+PEREK AFIEENRKKAAAR+RV+
Sbjct: 1  MVKKCYQVKGATGVLATAATVAGAVFAVKETIIDPREEKAAPIEENRKKAAARRVA 57

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 862

A DNA sequence (GBSx0914) was identified in *S.agalactiae* <SEQ ID 2609> which encodes the amino acid sequence <SEQ ID 2610>. This protein is predicted to be tRNA isopentenylpyrophosphate transferase (miaA). Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9897> which encodes amino acid sequence <SEQ ID 9898> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BA06085 GB:AP001515 tRNA isopentenylpyrophosphate transferase
[Bacillus halodurans]
Identities = 139/311 (44%), Positives = 200/311 (63%), Gaps = 21/311 (6%)

Query: 7  KIKLIAVVGPTAWGKTALGIETLAKTFNGRIISGDSQQVYQKLDIGTAKASKEEQEYVHH 66
      K  KLIA+VGPTAVGKT  + LAK  NGR+ISGDS  QVY+ +DIGTAK + EE +  HH
Sbjct: 2  KEKLVAVVGPTAWGKTKTSVMAKRLNGEIVSGDSMQVYRMDIGTAKITAEEMDGVPHH 61

```

Query: 67 LIDVREVNENYSVYDFVKEAKVAIDTIISKGKIPIIVGGTGLYLQSLFEGYHLSGVEVNC 126
 LID+++ +E+SV DF A I I +G+P +VGOTGLY+ ++ +LG E
 5 Sbjet: 62 LIDIKDFSESFSVADPQOLATPLITEIHERGRILPFLVGGTGLYVNAVTHQNLGDIRADE 121

Query: 127 TLMAYREKLE---SLSDDELFEKLT---EQSIIIPQVNRRAIRALELAKF----- 171
 YR +LE S + L +KL+ + + I N RR TRALE+ K
 10 Sbjet: 122 D---YRHELEAPVNSVYGQALHDLKISKIDPKAAALHPNRYREVIRALEIKLTKGTUTE 178

Query: 172 -GNDLQNSSEPYDVLILGNDNRQVLYDRINRRVLDLMDNGLLDEAKMLYD-NTPSVQAS 229
 + + SPV+++IGL +R VLYDRINRRVD M++ GL+DPAK LYD Q+
 15 Sbjet: 179 QARHESETPSPYNLMIGLIMERDVLDRINRRVDMVBSGLIDEAKLYDRGRDQGSV 238

Query: 230 KGIGYKELFPYFSKQIPLEAEVDLKNQNTFRFAKRLQITWFRNMVVEFIMVGRNTQQKI 289
 + IGVE++ Y + LREA+D LK+N+RR+AKRQLTWFRN+ NV + + + +KI
 15 Sbjet: 239 QAIQYKMYDYLDGNVTLEBAIDTLKNSRRYAKRQLTWFRNQNVTMPDMTDVDTKKI 298

Query: 290 KRKVSDFLSK 300
 ++ +F++ K
 20 Sbjet: 299 -MEIHNFIAK 308

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2611> which encodes the amino acid sequence <SEQ ID 2612>. Analysis of this protein sequence reveals the following:

Possible site: 27
 25 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 202/296 (68%), Positives = 250/296 (84%)

Query: 5 MRKIKLIVAVGPTAVGKTAIGLIEAKTPNGRIISGDSQQVYQKLDIGTAKASKEEQAY 64
 M KIK++ +VGPTAVGKTALGI LAK FNGEISGDSQQVY++LDIGTAKA++EEQE A
 35 Sbjet: 1 MTKIKIVIVGPTAVGKTALGISLAKAFNGEISGDSQQVYRQLDIGTAKATCEQEAAV 60

Query: 65 HHLIDVREVNENYSVYDFVKEAKVAIDTIISKGKIPIIVGGTGLYLQSLFEGYHLOGEVN 124
 HHLID+REV E+YS YDFV++A+ I I +S+GK+PIIVGGTGLYLQSLFEGYHLOGV+V
 40 Sbjet: 61 HHLIDIREVTESYAYDFVQDAQISIDVSRGKLPIIVGGTGLYLQSLFEGYHLOGQVD 120

Query: 125 OSTMAYREKLESLSDELFEKLTQSIIIPQVNRRAIRALELAKFGNDLQNSSEPYDV 184
 QE + AYR +LE L D DL+E+L +I I QVNRRAIRALELA+F ++L+N+E+ Y+
 45 Sbjet: 121 QSAVKAYRNELEQLDHDLYERLQVNNITTEQVNRRAIRALELAQFADELNBAETAYEP 180

Query: 185 LLIIGLNDNRQVLYDRINRRVLDLMDNGLLDEAKMLYDNPVQASKGIGYKELFPYFSKQ 244
 L+IGLNDNRQVLYDRIN+RV+ M+ +NGLL+RAKMLY++YP+VQAS+GIGYKELFPYF +
 50 Sbjet: 181 LIIGLNDNRQVLYDRINQVRNRMTEGKLLDEAKMLYEHYVYVQASRGIGYKELFPYFVGR 240

Query: 245 IPLEAEVDLKNQNTFRFAKRLQITWFRNMVVEFIMVGRNTQQKI 300
 + L EA D+LKNQNTFRFAKRLQITWFRNM V F + +Y Q + +V DFL K
 Sbjet: 241 MTLAASDQLKNQNTFRFAKRLQITWFRNMVAVSFITADVPQVHVRVDFLQK 296

55 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 863

A DNA sequence (GBSx0915) was identified in *S.agalactiae* <SEQ ID 2613> which encodes the amino acid sequence <SEQ ID 2614>. This protein is predicted to be hflX (hflX). Analysis of this protein
 60 sequence reveals the following:

-949-

Possible site: 35

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

- 5 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 10 >GP:BAB06081 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 182/406 (44%), Positives = 254/406 (61%), Gaps = 12/406 (2%)
- Query: 9 ERVILVGVLEQDT--ENFMSMERLASLAKTAGANVVVHHYQKRDKYDSKSFISGKLE 66
 ERV LV +L + E F3 S+KKL +L TA V++ QKR+ + ++IG GKLE
 15 Sbjct: 10 ERVFLVACQLPNMTDEQFEASLELEALT/LTAQGTVIDRLTQKREAIPTATYIGRKLDE 69
- Query: 67 IKAIVEADEIDTVVHNRLTPRQNSNLEASLGKVIDRMQLIIDIFAMRARSHEGKLQVH 126
 + +E E D V-VN L+ Q NL LGV+VIDR QLIIIDIFA RA+S BGLQV
 20 Sbjct: 70 LAIYNREQEADLVINGELSGSQVNLNRLGVRVIDRTQLIIDIFAGRAKSREGKLQVE 129
- Query: 127 LAQLKMYLPRLVGGGIMLSRQAGGIGSRGPGSQLENNRSIRHQISIDIERQLKIVKNNR 186
 LAQL Y+LVR+VGGG LSR GGIG+RGPGB++LE +RR IR ++DI++QLK K+R
 25 Sbjct: 130 LAQLNYLPRIVGGGQLSRGLGGIGTRGPGETKLETRHRIKRRADIDKQLHTVVRH 189
- Query: 187 ETVRRERDSTTFKIGLLGTYNAGKSTIMNVLTDKQYANELPATLDATTQIYVQNQF 246
 + R RR + TP-I L+GYNAGKST+N LT YE + LPATLD T++ L +
 30 Sbjct: 190 DRYRARRERNQTFRIALVGYTNAGKSTLINRLTASDSYEEDLLPATLDPMTRCRLSGM 249
- Query: 247 QVTLDTDTGFIQDLPTFLVAAFKSTLEESRAVOLLFHVIDASDPNHEEKKVMEILLDL 306
 +V L+DTVGFI LPT LVAAF+STLEE +H DLL HV+D S + H + V E+L L
 35 Sbjct: 250 EVLLSDTVGFIQLPTTLVAAFSTLEEKGADILLHVDRSSBQLQAHMETVSELHQL 309
- Query: 307 DMIDIPRLAIYNHMDVTEQLNATTP-----NVRIAAKKQGSKOLLRLVIDEIRHFDE 361
 ++ L +YNK D + N P ++A K+ LR++I + +F
 40 Sbjct: 310 EVDQSGMLVYNKAD---KNPLFIIPVHQQGISMSAHKREDIQLRQMRKTLVDLFTF 366
- Query: 362 FSRVHQNQAYKLYNDKIALLDITTYTFEEVE--NITGYISPKQW 405
 + + ++ KL L + ++ ++E+ E + GY+ P W
 45 Sbjct: 367 YVTELASDRGNKLAKRLRETIMTEMKWDREBCYQVGYVHPNHW 412

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2615> which encodes the amino acid sequence <SEQ ID 2616>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

- 50 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

- >GP:BAB06081 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 185/403 (45%), Positives = 246/403 (60%), Gaps = 6/403 (1%)
- 55 Query: 13 ERVILLGVLE--QTTEHFMSTELANLAKTAGVVKWMAFSQKREYDSKTFISGKLE 70
 ERV L+ +L T E F+ S+ EL L TA V+ +QRE + T+IG GKLE
 Sbjct: 10 ERVFLVACQLPNMTDEQFEASLELEALT/LTAQGTVIDRLTQKREAIPTATYIGRKLDE 69
- 60 Query: 71 IKAIVEADEIDAVIVHNRLTARQANLRAVLEVKVIDRMQLIIDIFAMRARSHEGKLQVH 130
 + +E E D VIVN L+ Q NL L V+VIDR QLIIIDIFA RA+S BGLQV
 Sbjct: 70 LAIYNREQEADLVINGELSGSQVNLNRLGVRVIDRTQLIIDIFAGRAKSREGKLQVE 129
- Query: 131 LAQLKMYLPRLVGGGIMLSRQAGGIGSRGPGSQLENNRSIRHQIADIIRQLTQVNNR 190

-950-

LAQL Y+LPR+VGQG LSR G3IG+RGPE+LE +RR IR ++ADI++QL K+R
 Sbjct: 130 LAQLNYLLPRIVGGQGLSRKLG3IGTRGPGETKLETRRHIRKRMADIKQLKHTVKIR 189

5 Query: 191 QTIRDRRVGSSTFKIGLIGYTNAGKSTIMNLLTDDSHYANELPATLDATTQQLYLNQF 250
 R RR + TP+I L+GYTNAGKST+N LT YE + LPATLD T++ L +
 Sbjct: 190 DRYKARRSRNQTFRIALVGYTNAGKSTILNRLTASDSYEDLLPATLDPTWKRLPLSGM 249

10 Query: 251 QATLTDTVGF IQDLPTSLVAAPKSTLEESKIVDLLHVIDASDPNHSBQEKVVLNKKEL 310
 + L+DTVGF1 LPT LVAAP+STLE K+ DLLHVD+D S + V LL +L
 Sbjct: 250 EVLSDTVGF INQLPTTLVAAPRSTLEEVKADLLHVVDRSSBQLQAHMETVSELLAQL 309

15 Query: 311 IMINIPRLAIYINKVDIASQ--FTATAPFNITISARSKDILARLIIQDQIRDFVPRFI 368
 ++ L +YNK D I +SA ++ LR++I + D F P+
 Sbjct: 310 EVDQSGMLVYINKADKPNLEIIFVHQNGIEMSAHKRSDIQRLAQMTIRLVDLPTPYVT 369

Query: 369 KYHQKAYKLYDLNRVALLDHYTFDQRIE--DISGYISPKQOM 409
 ++ D+ KL L R ++ +D++ E + GY+ P W
 Sbjct: 370 ELASDGNKLAKLRRTITMKNWDBRCYQVKGIVHFNHAW 412

20 An alignment of the GAS and GBS proteins is shown below.

Identities = 326/412 (79%), Positives = 375/412 (90%)

Query: 1 MIETKEQERVILVGVELQDTNPFMSMRELASLAKTAGANVNNHYQKRDKYDSKSPIG 60
 MIETK +QERVIL+GVELQ TB+P+MSM ELA+LAKTAG V+ + QKR++YDSK+FIG
 25 Sbjct: 5 MIETKQKQERVILGVELQTTTFHFDMSMETLANLAKTAGVKVNASFSQKREYKSTKPI 64

Query: 61 SGKLEIKAIWADEIDVTVVNNRLTPRCNNSLEAELGVKVIDRMQLILDIFAMRARGHE 120
 SGKL+BIKAIWADEID V+VNNRLT RGN+NLEA L KVIDRMQLILDIFAMRARGHE
 30 Sbjct: 65 SGKLDIKAIWADEIDAVIVNNRLTARONANLEAVLEVKVIDRMQLILDIFAMRARGHE 124

Query: 121 GKIQVHLAQLKYMFLRVLGGQIMLSRQAGGIGSRPGESQLELNRRSIRHQIDIERQL 180
 GKIQVHLAQLKYMFLRVLGGQIMLSRQAGGIGSRPGESQLELNRRSIRHQI+DIERQL
 35 Sbjct: 125 GKIQVHLAQLKYMFLRVLGGQIMLSRQAGGIGSRPGESQLELNRRSIRHQIADIERQL 184

Query: 181 IVEKNRRTVERRDVSTFTKIGLIGYTNAGKSTIMNLLTDDSHYANELPATLDATTQQL 240
 VEKRR+T+R+RRV S TFKLIGLIGYTNAGKSTIMN+LTDD YEANELPATLDATTQQL
 40 Sbjct: 185 QVEKNRQTIRDRKVSOTFTKIGLIGYTNAGKSTIMNLLTDDSHYANELPATLDATTQQL 244

Query: 241 YLQNQFQVTLDTVGF IQDLPTSLVAAPKSTLEESKIVDLLHVIDASDPNHSBQEKVVL 300
 YL+NQFQ TLDTVGF IQDLPTSLVAAPKSTLEES+VDLL HVIDASDPNH E ERV+
 45 Sbjct: 245 YLQNQFQATLDTVGF IQDLPTSLVAAPKSTLEESKIVDLLHVIDASDPNHSBQEKVVL 304

Query: 301 EILKDLWDIDIPRLAIYINKMDVTEQLNATTFFNVRJAARKQGSKOLLRLVIDEIRHFD 360
 +LK+LDM++IPRLAIYINK+D+ EQ AT FNN+RI+A+ + SK LERLL+D+IR F
 45 Sbjct: 305 NLKLELOMLNIPRLAIYINKVDIASQPTATAPFNIRISARSKDILARLIIQDQIRDFV 364

Query: 361 EFSIRVHQVQAYKLYDLNRVALLDHYTFDQRIEIDISGYISPKQWLLDFTYE 412
 F I+VHQ++AYKLYDLN+ALLD YTF+E E+I+GVISPKQ+W+L+FY+
 50 Sbjct: 365 PFRKIVHQKAYKLYDLNRVALLDHYTFDQRIEIDISGYISPKQWLLDFTYE 416

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 864

55 A DNA sequence (GBSx0916) was identified in *S.galactiae* <SEQ ID 2617> which encodes the amino acid sequence <SEQ ID 2618>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence

60 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2044(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2619> which encodes the amino acid sequence <SEQ ID 2620>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 40
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3436 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 124/209 (59%), Positives = 150/209 (71%)

15      Query: 1 MIDYIDLALTYGGFTSLDKVYLEKKLOGLSKQRLDFTITPPPSVINAYFAEIQKQSGPEA 60
      M +YIDIA TYGGFTSLD YL L L+ QQ+L FITPPPSVINAYFARIYQKQ P+A
      Sbjct: 5 MNNYIDLAKTYGGFTSLDTNYLNHLASLTDQCKLAFITPPPSVINAYFAEIQKQSPQA 64

20      Query: 61 ATDYFFDLKALGLFPKHLSDFEKKPFIRINLGGKSPGFAYLNDCEASVPSEKVEVITP 120
      ATDYFF+LKKALGLF SF+EEKPF+RLNLSGK+-GFAY NDCE A VPSE E P
      Sbjct: 65 ATDYFFNLKALGLFTDQPSFEKKPFVRLNLSGKAYGFAYNDQCEVALVPSEKSPKPF 124

25      Query: 121 QLLLEIAQIFPQYKYVRDSGIRMAKIDFDETESQNITPETSLLGNVLQKLDIITKTSF 180
      +L E+ QIFPQY VY D+ ++M F++ E ++ITP+ +LL + +L I + F
      Sbjct: 125 ELFFELIQTIFPQYMYVEDKQGLQKQAKQFQEGCEDITFDOTLLSKIYRLANGITMLKGF 184

      Query: 181 NQBELLELVKTKSGKYYSSQGRSEVIY 209
      N EEL L +T SG+ YY RE +IYI
30      Sbjct: 185 NVEELNALSGTFSGQKYDFAQREFMIYI 213

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 865

35 A DNA sequence (GBSx0917) was identified in *S.galactiae* <SEQ ID 2621> which encodes the amino acid sequence <SEQ ID 2622>. Analysis of this protein sequence reveals the following:

```

      Possible site: 16
      >>> Seems to have no N-terminal signal sequence

40      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1060 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

45 A related GBS nucleic acid sequence <SEQ ID 9895> which encodes amino acid sequence <SEQ ID 9896> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

      >GP:CAB14316 GB:299116 similar to hypothetical proteins [Bacillus subtilis]
      Identities = 156/309 (50%), Positives = 210/309 (67%), Gaps = 5/309 (1%)

50      Query: 1 MEIQFLTQAGQPAKARNVSSVLKLLDEINVMFDCQEGSTQKQILETTIKPRKVKKIF 60
      ME+ FLTQAG PAKARNV+S+ LKLL+E VM+FDQGE TQ QIL TTIKPRK++KIF
      Sbjct: 1 MELLFLTQAGTIPAKARNVTVALKLLERSSVLMFDQGEATQHQILHTTIKPRKIEKIF 60

55      Query: 61 ITHMGCDHVPLGFLGFLSSRAFOANESQITLDIYGPWGKISFVMTALRTSQRPLPYRIHFH 120
      ITHMGCDHV+GLPLG L SR+FO E++ L +YGP GIK+F+ T+L + + L Y +

```

-952-

Sbjct: 61 ITHHGDEHVGFLGFLSSRFQGGDE--LVYGPNGIKAFIETSLAVTHTLTYPLAIQ 118
 Query: 121 EFDSSSLGKIMETDKFTVYAEKLDHTIFCMGYRVVQKDLBGLTDLBALKLAGVPPGFLPG 180
 E +E G + E D + F V A + H + GYRV +KD+ G+L A+ LK +P GP+
 Sbjct: 119 EIEE---GIVFEDDQPIVTVAVSVIHGVEARFYRVQKDVPGSLKADVLKMMIIPPGFVYQ 175
 Query: 181 KVNGENVLTLEDGREIIAKDYISEPKKGVITILGDTKTRTASIRLALGADVLVHSESTYG 240
 K+K GE VFLRDR I D++ PKKG++ GDTR +D IA DVLVHE+T+
 Sbjct: 176 KIKKGFTVLTLEDGRILINGNDPLFEPKPKGRSVFSGDTRVSDKLAKLRDQDVLVHEATFA 235
 Query: 241 KQDERIAKSHGHSNMQAADIAKQANAKRLLLNHVSARFMGRDCWQMEEDAKTIPFNTHL 300
 K D ++A + HST QAA AK+A AK+L+L H+SAR+ G ++++A +F N+
 Sbjct: 236 KEDRKLAIDYHISTTEQAATVAKERAKQLIATISARYQGDASLRLEKRAVDVFPNSVA 295
 Query: 301 VRDLEEVGI 309
 D EV +
 Sbjct: 296 AYDFLEVVN 304

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2623> which encodes the amino acid
 sequence <SEQ ID 2624>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2352 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 253/307 (82%), Positives = 285/307 (92%)
 Query: 1 MEIQFLGTGAGQPAKARNVSSILKLLDEINEVMMFDCBGTQRCILETTIKPRKIKIP 60
 ME+QFLGTGAGQPAK RNVSSL LKLLDEINEVMMFDCBGTQRCILETTIKPRK++KIP
 Sbjct: 1 MEIQFLGTGAGQPAKARNVSSILKLLDEINEVMMFDCBGTQRCILETTIKPRKIKIP 60
 Query: 61 ITHHGDEHVGFLGFLSSRFQGGDE--LVYGPNGIKAFIETSLAVTHTLTYPLAIQ 120
 ITH+HGDH+FLGFLSSR+FGA+BEQTDLDIYGP+GIK++V+T+L+ SG+R+PY+IHFH
 Sbjct: 61 ITHLHGDEHVGFLGFLSSRFQGGDE--LVYGPNGIKAFIETSLAVTHTLTYPLAIQ 120
 Query: 121 EFDSSSLGKIMETDKFTVYAEKLDHTIFCMGYRVVQKDLBGLTDLBALKLAGVPPGFLPG 180
 EFD+ SLGKIMETDKF VYAE+L HTIFCMGYRVVQKDLBGLTDLBALK AGVPPGFLPG
 Sbjct: 121 EFDKSLGKIMETDKFEVYAEKLAHTIFCMGYRVVQKDLBGLTDLBALKAGVPPGFLPG 180
 Query: 181 KVNGENVLTLEDGREIIAKDYISEPKKGVITILGDTKTRTASIRLALGADVLVHSESTYG 240
 K+KNG++V LEDGR I AKDYIS PKKG+ITI+GDTRT AS++LA ADVLVHSESTYG
 Sbjct: 181 KIKNGQDVLLEDGRILCAKDYISAPKKGKITIIGDTRTASVSKIAKDADVLVHSESTYG 240
 Query: 241 KQDERIAKSHGHSNMQAADIAKQANAKRLLLNHVSARFMGRDCWQMEEDAKTIPFNTHL 300
 KQDERIA++HGHSTNMQA IA +A AKRLLLNHVSARFMGRDCWQMEEDAKTIPFNTHL
 Sbjct: 241 KQDERIAKSHGHSNMQAQIAHAGAKRLLLNHVSARFLGRDCRQMEEDAKTIPFNTHL 300
 Query: 301 VRDLEEV 309
 V+DLEEV
 Sbjct: 301 VQDLEEV 307

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 866

A DNA sequence (GBSx0918) was identified in *S.agalactiae* <SEQ ID 2625> which encodes the amino acid sequence <SEQ ID 2626>. This protein is predicted to be similar to ketoacyl reductase. Analysis of this protein sequence reveals the following:

```

5   Possible site: 17
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15 >GP:CAB14310 GB:Z99116 similar to ketoacyl reductase [Bacillus subtilis]
   Identities = 100/253 (39%), Positives = 152/253 (59%), Gaps = 2/253 (0%)

   Query: 3   RTILITGASGGLAQAIINQLPQDD-HLIVTGRSREKLEKLYGHRPNILCLSLDITN-DNA 60
           + I ITGASGGL + I + H++++ R ++L ++ K +I D
   Sbjct: 7   KRIWITGASGGLGERIAYLCAABGAVLVLSARREDRLTEIKRKITEBWSQCRIFPLDVG 66

20   Query: 61  VTNMIEKIYGERGQIDILINNAFGSGFKFMDYSDERVKDMFVAVNTFATMSIARQIGHKM 120
           I ++ + G ID+LINNAFGF + D + ++K MF VN F ++ + + +M
   Sbjct: 67  RLEDIARVRDQIGSIDVLINNAFGFIFETVLDSTLDDMKAMPDVNVFGLIACKAVLPQM 126

25   Query: 121 SLVKSGHIVNTIASMAGLIATSKASVYGASKFAVVGFSNALRLELAKKVVYVTSVNPGPQK 180
           K GHI+NIAS AG IAT K+S+Y A+K AV+G+SNALR+EL+ +YVT+VNPGP+
   Sbjct: 127 LEQKKGHIINIASQAGKIATPKSSLYSATKHAVLGYSNALRMELSGTGIYVTVNPGPIQ 186

30   Query: 181 TGFFAQADPSPGYLASIGRFALTPEKVKKVVSLGKNKRELILPFLIAPANKYISLFPK 240
           T FF+ AD GUY +GR+ L F+ V+ ++ + + KRB+NLP ++ K Y LFP
   Sbjct: 187 TDPFSLADKGGYAIQVGRWMLDPPDVAQITAAITFKKREINLPRLNAGTKLYQLFPA 246

   Query: 241 TADYPARKVFNFK 253
           + A + K
35   Sbjct: 247 LVEKLAGRAIMKK 259

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2627> which encodes the amino acid sequence <SEQ ID 2628>. Analysis of this protein sequence reveals the following:

```

40   Possible site: 18
   >>> Seems to have a cleavable N-term signal seq.

   ----- Final Results -----
45          bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

50 >GP:BA805225 GB:AP001512 oxidoreductase [Bacillus halodurans]
   Identities = 107/259 (41%), Positives = 156/259 (59%), Gaps = 5/259 (1%)

   Query: 1   NAQRRIIVITGASGGLAQAIKQLPKEDSLI-LIGRNKRELKHICYQHI----DNKCECLLD 55
           M ++ I ITGAS GL + + E++++ L R++RLE+ + + +D
   Sbjct: 1   MRKKTIFITGASSGLGRQLADFSMEETVLCLFARSQHLSENVRQIVVENGGGEAHITYPD 60

55   Query: 56  ITNFAIEKMWAIQIYQYGRIDVLINNAAGYGAFKGFRRPSAQRIADMPQVNTLASHPAC 115
           + +P +I++ R+ G +HVLINNAOYG F + F + E MF+VN +
   Sbjct: 61  LADPQSIDRSFARAI SAVGVVDVLINNAAGYGVFEFCDSQMDENRMFRVNVFGLMRATA 120

60   Query: 116 LIGQMAEQSQGHILINIVEMAGLIASAKSSIYSATKPALIGFSNALRLELADKGVVYVTV 175
           + M EQG GH+INI S AG IA+AKS+IYSATK A++GF+N+LR+EL G++V+ V

```

-954-

Sbjct: 121 AVLPTMREQSGGHIINIASQAGKIATAKSAIYSATKHAVLGFTNSLMELKOTGIHVS AV 180

Query: 176 NPGPIATKFPDQADPSGHYLESVGKFTLCQNOVAKRLVSIIGKKNRELNLFFSLAVTHQF 235
NPGPI T PFDQAD G Y V + L P V + + + V + K KRELNLFF + + +

5 Sbjct: 181 NPGPIQTFFPDQADKSGAYTSKVQRIMLDFEDVSRKIVQLTKKPKRELNLFPW@NIGATA 240

Query: 236 YTLFPEKLSDYLRKVFENK 254
Y + P+L + LA K F K

10 Sbjct: 241 YQVAFRLLELLAGKQPRQK 259

An alignment of the GAS and GBS proteins is shown below.

Identities = 155/251 (61%), Positives = 200/251 (78%)

Query: 3 RTILITGASGGLAQAIINQLPODDHLIVTGRSREKLEKLYGKRPMTLCSLDITNDNAV 62
R I+ITGASGGLAQAI+ QLP++D LI+ GR++E+LE Y N CL LDITN A+
15 Sbjct: 4 RTIVITGASGGLAQAIQVQLPKEDSLILLGRNKRERLEHCYHIDNKECELDITNFVAE 63

Query: 63 NMIEKIYGFEGQIDILINNAGFGSGFKEPFDWSDSEVKKMFVNTVATMSIARQIGKHSL 122
M+ +IY +G+ID+LINNAG+G+FK F ++S +E+ DMF VNT A++ A IG KH+
20 Sbjct: 64 KMVAQIYQRYGRIDVILINNAGYGAFKFEPEPSAQELDMFQVNTLASIHFACLIGQQAEE 123

Query: 123 VKSGHIVNTIASMAGLIATSKASVYGAKFAVVGFSNALRLLEAKNVVYTSVNRGPIKTS 182
GH+INI SMAGLIA++K+S+Y A+KFA++GFSNALRLLEA+K VYTV+VNRGPI T
25 Sbjct: 124 QSQGHLINIVSMAGLIASAKSSIYSATKTFALIGFSNALRLLEADNSGVYTVTVNRGPIATK 183

Query: 183 FFAQADPSGDYLASIGRFALTPEKYSKCVVSILOKHKRELNLFFILAFAHKYSLFPKTA 242
FF QADPSG YL S+G+F L F +V+K++VSI+GKHKRELNLFF LA H++Y+LFPK +
30 Sbjct: 184 FFDQADESGHYLESVGKFTLOFNQVAGLVSIIGKKNRELNLFFSLAVTHQFYTLPFKLS 243

Query: 243 DYFARKVFNKY 253
DY ARKVFENK
Sbjct: 244 DYLRKVFENK 254

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 867

A DNA sequence (GBSx0919) was identified in *S. agalactiae* <SEQ ID 2629> which encodes the amino acid sequence <SEQ ID 2630>. This protein is predicted to be single-stranded-DNA-specific exonuclease (recJ). Analysis of this protein sequence reveals the following:

Possible site: 31
>>> Seems to have no N-terminal signal signal
INTEGRAL Likelihood = -0.16 Transmembrane 197 - 213 (197 - 213)

---- Final Results ----
bacterial membrane --- Certainty=0.1065 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14721 GB:Z99118 similar to single-strand DNA-specific
exonuclease [Bacillus subtilis]
Identities = 276/772 (35%), Positives = 447/772 (57%), Gaps = 45/772 (5%)

Query: 1 MISAKYSWVLNCKIDAGFFEASKKE-KISAVASLIYSRGIKTSARLHHFLQTNLENLH 59
M++K W + Q+PD ++ ++ I+ VASL+ RG T+ FL T ++
55 Sbjct: 1 MLASGRWBI--QRFDQVKSLTEQLHITFLVASLLVGRGFTTASARLFLHITDADPY 58

Query: 60 DPLYLNDNDKAVNRIRRAIKNNETITLVYGDYDADGMSASIMKEALDMMGARVQVLPNR 119
DE+ + M +A +RI++AL E I++YGDYDADG+TS S+M L +A+V Y+P+R
60 Sbjct: 59 DPEFMGKMKRAADRIKQISQSEKIMYGDYDADGVTS+VSMHLTKLSAQVDFYIPDR 118

-955-

Query: 120 PTDGYGNPQSVYKYPFIEQQVQVSLITVDMGVAGHEAITVAQNGQVDDVVDHHSMPADLP 179
 P +GXYGN+ ++ I+++ SLITVVD G+A A+ G+DV++TDHH ++LP
 Sbjct: 119 FKEGYGNPQAFPS-IXERGFSLITVDTGIAAVHEAKVAKELGLDVIITDHHSPGPELF 177

5 Query: 180 CAYAIHPEHPDANYPPFYLAGQGVAFKACALLETIPTFMDLDVAIGTIAADMVSLITEN 239
 AI+HP+ P YPF LAG GVAFK+A ALL +P E+LDL AIGTIAD+V L DEN
 Sbjct: 178 DVRAIVHPKQPGCTYPPFKLAGGVAFKALHALGELPDRLLDLAAIGTIAIDLVPIDHEN 237

10 Query: 240 RIMVYAGLEVMDKSERIGLQELISLSNIDLKTNNEETIGFKIAPQIALNALGRIDNDPATE 299
 R++ GLS ++ +R+GL+ELI LS D+ NEEY+GT++N+I+N+GR++ ++P+
 Sbjct: 238 RLIIATLGLERLRTNRLSLGLKILSGGDIGRANKEITVGQLAPRINAVGRIGEDQADPAVH 297

15 Query: 300 LITGTFDEESQAIAQIMIDQKNEERKRIQVTFIDQAMQMIDQ---TKPQVVLAKENHPGV 356
 LL D E++ +A IDQ N+ER+++V +D+A++M++Q + V+A K W+RGV
 Sbjct: 298 LMMSDSFEAEELAAEIDQLNKERQKMMVSKMTDRAIEMVBCQQLDQCAIVAKAGNPGV 357

20 Query: 357 LGIVAGRIIEFTGQPVIVLNI--KDIAGKSARSVEALDIFQAFDQHRLEFLAPGCHSGA 414
 +GIVA +++++R +P IVL I E GIAGKGARS+ ++F++ +R++ PGGH A
 Sbjct: 358 VGVIVASKLVDRFYRPAIVLGIIEKGIAGKSARSIRGFNLPESLSECRDLTHPFGCHPMA 417

25 Query: 415 AGMTLRESKQVGLDQVLCYIISKKQLMSQKKTLLINDSLRFDRLSLDTVRDFEKLAPFG 474
 AGMTL+ V DL L + + +D +++++++ + L+PFG
 Sbjct: 418 AGMTLKAEDVPDLRSRLNEIADNTLTERDFIPVQEVDLVCGVEDITVESIAEMMLSPFG 477

30 Query: 475 MNNKKPFLNDFPK/SQARVMQNGAHLKLLKLEQDQALDLVAFNMMSQLQEFQOACHLE 534
 M N KP L+++ + R +G N H+K+ + + LD V FN G + +
 Sbjct: 478 MLNPKPHVLVENAVLEEDVRKIGANKTHVMRTIRNESSQLDCVGNFKGLQEGIVPQSRIS 537

35 Query: 535 LAVTLVNVNMKGATTLQMLEDAFVDGIQLDFIRSK-----ASSLEPG----- 577
 + +S+N+NN QM++DA V QLPDR K S+LP
 Sbjct: 538 IYGEISINENNRKKPKQMIDKAAVSEWQLDFLRGRTWEDTVSALPSAKRAIVSPKDS 597

40 Query: 578 -----VPILSQECSKE-----VILLTVPHPQELKMQTKQKQFDALYFNK 618
 V ++S ++Q+K ++LL P L ++ +GK + IYF
 Sbjct: 598 TTLQTEDLRREVVISSKQAKAFDQDLDGAYIVLLDPPPSLMDLARLEGGKAPERYITIF 657

45 Query: 619 RIFKNYFISGYGTRDQFASLKYTIYQFPEFDVRYKLKELSXYLHIDPILLIMIQIFEEEL 678
 +++F+S + RD F Y + + FDV+ EL+ + + M ++F +L
 Sbjct: 658 LNHEDHFLSTPPARDHPKMYAFLLKRGAFQVKKHGSSELAHGKNSVETINFMTKVFYFDL 717

Query: 679 HFVTITGIMTVNKEAKRDISQSQIYQELKETVKPQELMALGTPKEIYDPM 730
 FV I G+++V A+KRD+++SQ YQ ++ ++ + + +E+ +++
 Sbjct: 718 GFVKIENGVLSSVSGAKRDLTDSQYQAKQQLMELDQLQNLAYSAREELKEWL 769

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2631> which encodes the amino acid sequence <SEQ ID 2632>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

50	INTEGRAL	Likelihood = -0.16	Transmembrane	220 - 236 (220 - 236)
	INTEGRAL	Likelihood = -0.11	Transmembrane	667 - 683 (667 - 683)

----- Final Results -----

55	bacterial membrane	---	Certainty=0.1065 (Affirmative)	< succ>
	bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
	bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 474/731 (64%), Positives = 594/731 (80%)

60 Query: 1 MISAKYSWVIANQKPDAGFPFASKKEIKSEAVASLIYSRGIKTSARLHHFIQTNLNLDH 60
 MI +KYSW + ++KPD GFF+ +K + ++ A LIY RGI+T L FL +L LHD
 Sbjct: 1 MIRSKEWIKKIDKPDGQFKAETKGLTQTAQLIYDRGIRTEALDEFULADLSCQEDH 60

65 Query: 61 PYLLNDMDKAVNRIRRAITENNNTILVYGDYDADGMSASIMKSAIDMDGAEQVYLPNRF 120

-956-

```

      PYLL+DM KAV RIR+AIR E IL+YGDDADGMPTEASI+KE LDMMGAR VYLPNRF
Sbjct: 61 PYLLHDMKAVPRIRQAIEBGERILLYGDDADGMPTEASIVKETLDMMGARELVYLPNRF 120

5   Query: 121 TDGYGPNQSVYKYFIBQ+VSLIITVDNMGVAGHBAITYAQMSQVNVVTDHHSMPADLPC 180
      TDGYGPNQSVYKYFIBQ+ VSLIITVDNMGVAGHBAI YAQ Q VDV+VTDHHS+P +LP
Sbjct: 121 TDGYGPNQSVYKYFIBQ+VSLIITVDNMGVAGHBAITYAQMSQVNVVTDHHSMPADLPC 180

10  Query: 181 AYAIHHEHPDANYPFPYLAGQVAFKACALLEITFTPEMLDLVAIGTIADMSVLSIDENE 240
      A+AIHHEHPDA+YFP +LAGQVAFK+ A ALLE++FT+ LDLVAIGTIADMSVLSIT ENR
Sbjct: 181 AFAIHHHEHPDADYFPKHLAGQVAFKALATLLESPLTDCLDLVAIGTIADMSVLSIGENE 240

15  Query: 241 IMVAGLEVMDSERIGLQELISLMSIDLKTLNEETIGFKIAPQNALGRLLDOPNEAIEL 300
      ++VK GL ++K +ER+GLQEL+SLS IDL+ NE+ IGF+IAPQNALGRLLDOPNEAIEL
Sbjct: 241 VLVRGLAMLEKTERVGLQELMSLSFDLHFNHDAIGFQIAPQNALGRLLDOPNEAIEL 300

20  Query: 301 LTGFDDERSQAIQMIDQKNSEKKEITQTFIDQMMLDQTPKPVUAKENWHGVLGIY 360
      LTGFDD+R+QAIA MI +KNEERK +VQ IFDQM N+D KPVQVIR+ WHRGVLGIY
Sbjct: 301 LTGFDDQRAQAIALMIKKKNEERKALVQDIFDQMAMAVDPKPVQVLAQAGHPVGLGIY 360

25  Query: 361 AGRILERTGQPVIVLNIEDGIAGKSARSVEALDIPQAFDQHRELFIAPGGHSGAAGMTLE 420
      AGRIR+R GQ V+VL I++G AKGSARS+RA++IFA+ + RELF APGGH+GAAGMTL
Sbjct: 361 AGRIMETIGQTVVLTITINGFAKGSARSLEAINIFEALMGKRELFTAPGGHGAAGMTLE 420

30  Query: 421 ESKVGDLISVCLDYISKKQLDMSQKTLTIDSEIRFDELSLDTVRDFEKLAPFGMNKKR 480
      + LS LC ++ ++ LD + K TLTID L D+LSLD ++ +KLAP+GMD+KP
Sbjct: 421 VDNLEALSDFLQGVIERGLDQTAKNTLTIDERSLDDLSLDILKSLDKIAPYGMCHQKP 480

35  Query: 481 VFLLKDFKVSQARVMGQNGAHLKKIKESDQALDILVAFNMGSLQCFQCAQHLRLAVTIS 540
      VF +KD +VSCAR +GQ+ +HLK K+ Q + D+AF GSLQCF+QA LELAVTIS
Sbjct: 481 VPFVKDIRVSQARTIGDQSHLKFVKSQKASFDVLAFGQSSQLCFQFQATGELAVTIS 540

40  Query: 541 VNQWNGJATTQLMLEARDVVDGIGLFDIRKASSLEPHGVPILSQEQSKVILLVTVDEPQ 600
      VN WNG T+LQ ML DARVDG+QL D+R+K + +P G+P + ++ ++ ++ +E+ +
Sbjct: 541 VNHNGNTSLQFMELVARDVGVQLLDELTKTAKVPESGPTIREDPNARVILINDIPEDFK 600

45  Query: 601 ELKQMTGKQFDIAIFYKNEIPKNIYFISGQYTRDQFASLYKTIYQFPEFVRYKLKLSY 660
      + K FDAIFYKNI++ Y+++G+G+R+QFA LYKTIYQFPEFD+R+KL ELS Y
Sbjct: 601 TWRNQFVHKDFDAIFYKNIKHEFYLLTGFSREQFAKLYKTIYQFPEFDLRKHLTLESHY 660

50  Query: 661 LHIPDILLIKNIQIFELHFVTITEGIMTVNKEAKRDISHSITYQLKELTKVQPSLMAL 720
      L+I +LLIK+IQIFEL FVTI +G+MTVN +A+R+ISHS IYQ+LKE VKPOE+NAL
Sbjct: 661 LNIEKLLILLIKIQIFELSFVITIDGLMTVMNQAKREISESHITYQDLKELVKPOEINAL 720

Query: 721 GTPKRIYDPMN 731
      +PKS+YD+++
Sbjct: 721 ASPKSMYDYLW 731

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 868

A DNA sequence (GBSx0920) was identified in *S. agalactiae* <SEQ ID 2633> which encodes the amino acid sequence <SEQ ID 2634>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence

```

55

```

----- Final Results -----

```

```

      bacterial cytoplasm --- Certainty=0.4114(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

60

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 869

- 5 A DNA sequence (GBSx0921) was identified in *S.agalactiae* <SEQ ID 2635> which encodes the amino acid sequence <SEQ ID 2636>. Analysis of this protein sequence reveals the following:

```

Possible site: 42
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -5.10 Transmembrane 15 - 31 ( 14 - 33)

----- Final Results -----
bacterial membrane --- Certainty=0.3039 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AA88584 GB:M18954 fructosyltransferase [Streptococcus mutans]
Identities = 67/219 (30%), Positives = 106/219 (47%), Gaps = 31/219 (14%)

Query: 1 MRFIVRKMYYKKGKFWVAGIVT-ILGGSAILGQDVKAQAEAVTSTISEKTDSQTISD 59
M VRKMYKKGKFWVVA I T +L G + V+A++A + T SE +SQ +
Sbjct: 1 METKVRKMYKKGKFWVATITMTLGTGL--SSVQADEANS-TQVSSRLAERSQVQEN 57

Query: 60 TSKLTLPVNSSEAMONSAEFLIKTAFATSVSNPREIAATPVKTFDASSKVVVKASTAEH 119
T+ SS A +N A KT + S+NP AA V+ D ++KV+ + E
Sbjct: 58 TTA-----SSSAENQA---KTEVQETPSTNP---AAATVENTDQTKVITDGAIVES 104

Query: 120 SANQTN---SNVQVANDSEVITQDN-----STKQLPTVITYSAHVQDIGN---QKSVQ 166
A++T + V + A + + Q N +TK+ T + + G +K
Sbjct: 105 KASKTKDQAATVTKTAASTPEVGTQNEKD KAKATKEADITTPKNTIDEYGLTEQARKIAT 164

Query: 167 NATVSGTVGQEKQVEAIKLSIKAPEGITG-KLSYKTYK 204
A ++ + +KQVEA+ + TG +++Y+ + K
Sbjct: 165 EAGINLSSLTQKQVEALNKVKLTSDAQTHQNTYQEFDK 203

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

- 40 A related GBS gene <SEQ ID 8677> and protein <SEQ ID 8678> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop Possible site: -1 Crend: 5
McG: Discrim Score: 9.08
GVH: Signal Score (-7.5): -3.94
Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
ALOM program count: 1 value: -5.10 threshold: 0.0
INTEGRAL Likelihood = -5.10 Transmembrane 7 - 23 ( 6 - 25)
PERIPHERAL Likelihood = 4.03 694
modified ALOM score: 1.52

*** Reasoning Step: 3

----- Final Results -----
bacterial membrane --- Certainty=0.3039 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

31.1/52.1% over 749aa
5      EGAD[14681] levansucrase precursor Insert characterized
      SP[P11701|SACB_STKMV LEVANSUCRASE PRECURSOR (EC 2.4.1.10) (BETA-D-FRUCTOPURANOSYL
      TRANSFERASE) (SUCROSE
      6-FRUCTOSYL TRANSFERASE). Edit characterized
10     GP[153636|gb|AAA88584.1||M18954 fructosyltransferase Insert characterized
      FIR[B28551|B28551 levansucrase (EC 2.4.1.10) precursor - (strain GS-5) Insert
      characterized
      ORF02172(295 - 1731 of 3138)
15     EGAD[14681|14686(7 - 756 of 797) levansucrase precursor {Streptococcus mutans}
      SP[P11701|SACB_STKMV LEVANSUCRASE PRECURSOR (EC 2.4.1.10) (BETA-D-FRUCTOPURANOSYL
      TRANSFERASE) (SUCROSE 6-FRUCTOSYL TRANSFERASE). GP[153636|gb|AAA88584.1||M18954
      fructosyltransferase {Streptococcus mutans} FIR[B28551|B28551 levansucrase (EC 2.4.1.10)
      precursor - Streptococcus mutans (strain GS-5)
20     %Match = 2.9
      %Identity = 31.1 %Similarity = 52.1
      Matches = 83 Mismatches = 115 Conservative Sub.s = 56
      132      162      192      222      252      282      312      342
25     LFERLENGSQYQ*PYQH*YQ*RHNHHQYLVQ*ERVQQLIGRAPCL*PQFTVSYXXQN*LXYR*KMYKHKKFWVAGIV
      : ||||| |||||
      METKVRKKMYKLGKFWVAGIT
      10      20
30     372      402      432      462      492      522      552      582
      TLIGGSALAQDVABQAEAVTSTISEKTDSQTISDTSKLTLVNSSEAMNSAEPLIKATPATSVSSNPREIAATPVK
      | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
      TAM---LITGIGLSVQADEANST-----QVSSLEAERSQVQENTPSSAAENQARTVEQETPSTN---AAATVE
      30      40      50      60      70      80
35     612      642      663      693      705      735      783
      TFDASSKVVYKASTAEHSANCTNSN---VNQVANDSEVITQCN-----STKQLPTVTYSAHVQDIGN---QKSDVNAI
      | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
      NTDQTKTVITDAVNESKASKTKDQATVTKTAASTPEVQCTNEKDKAKATKEADITTPKNTIDRYGLTEQAKIATEAG
      100      110      120      130      140      150      160
40     813      834
      VSGTVQGEKQVEA---IKLSIKAPES-----ITGKLSYKTY
      : : ||||| : : || : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
45     INLSSTQKQVEALNKVKLTSDAQTGHQMTYQRFDKIAQTLLAQDE---VGTLDITAYLPGENDGYIDNIVIGVGYGLKPH
      180      190      200      210      660      670
50     912      942      972      1002      1032      1062      1092      1122
      VKQSGMQPQSVESQSGVSGTVGQSRPEALSLIMLDMNLKLYDVYRVHVQDIGNMAWAKNGAYATGLMSKELAEATVEKFT
      || : || : |
      TFSQ-YQPTV-----
      1152      1182      1209      1239      1269      1290      1320      1350
      LKQSGVLPTTIPKBERFVLYNQYKV-GQNGQGNKLEGGMAGTLGSKALDG---VKFTLSLKGYDILYRTHVQDNGWG
      | : | : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
55     -----FSTPTITDDIISFEVSDGHLVLPKPVVANDSAGRIDQSRNCGSLNVAFVSA-----
      690      700      710      720      730      740
      1641      1671      1701      1731      1761      1791      1821      1851
60     EI---SYQYTLQDGNKPTVLGSLQGLSGLSKSIKAIKINLGTALGNIEYRTPLNGSGNQTVVNSGREGVNPNSQQ
      : : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
      -----GGNISLVKPSQKSNNTKTKKAHHVSTKKKQKNSFFAALLALSPAFVSGIF
      750      760      770      780      790

```

SEQ ID 8678 (GBS243) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 57 (lane 7; MW 94kDa).

GBS243-His was purified as shown in Figure 208, lane 10.

Example 870

A DNA sequence (GBSx0922) was identified in *S. agalactiae* <SEQ ID 2637> which encodes the amino acid sequence <SEQ ID 2638>. This protein is predicted to be adenine phosphoribosyltransferase (apt).

5 Analysis of this protein sequence reveals the following:

```

Possible site: 59
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -1.86    Transmembrane    61 - 77 ( 59 - 77)
    INTEGRAL    Likelihood = -0.64    Transmembrane    137 - 153 ( 137 - 153)

----- Final Results -----
    bacterial membrane --- Certainty=0.1744(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
  
```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AA046040 GB:U86377 adenine phosphoribosyltransferase; Apt
[Bacillus subtilis]
Identities = 110/170 (64%), Positives = 135/170 (78%)

Query: 1 MDLNIYIASIENYPOGIGITFRDISPLNADGKAYSIVREIVQYAADKIDIMVGPPEARGE 60
      NDL Y+ +Y+PG+ F+DI+ LM G Y Y A +IV+YA +K ID++VGPEARGE
Sbjct: 1 MDLAKQVVTIVDPYKQGVQPKDITTLMDKGDVYRATDQIVETAKKQIDLVGPEARGE 60

Query: 61 IVGCFVAYALGIGFAPVRKPKGLPREVISADYEKEYGLDITLTHADAIKPGQRLVLDL 120
      I+GCFVAYALG+GFAPVRK GKLPREVI DY EYG D LT+H DAIKPGQRLV DDL
Sbjct: 61 IIQCFVAYALGVGFAPVRKPKGLPREVIKVDYGLYKQDVLTIHKDAIKPGQRLVITDDL 120

Query: 121 LATGGTIVKNTIEMIEKLGCVVAGCAFLVRLXGLANGKALIEGVDTIKVMNF 170
      LATGGT+ATI+++E+LGGVVG AFL+EL L+GR +E YD IM +
Sbjct: 121 LATGGTIENTIKLVERLGGVVGAGIAFLIELSYLDGRNKLREVDILTLMKY 170
  
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2639> which encodes the amino acid sequence <SEQ ID 2640>. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
    bacterial outside --- Certainty= 0.300(Affirmative) < succ>
    bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
  
```

The protein has homology with the following sequences in the databases:

```

[GB:Z99120 similar to opine catabolism [Bacillus sub... 231 1e-59]

>GP:CAH15253 GB:Z99120 similar to opine catabolism [Bacillus subtilis]

Score = 231 bits (583), Expect = 1e-59
Identities = 138/363 (38%), Positives = 212/363 (58%), Gaps = 11/363 (3%)

Query: 5 IIGAGIVGSTAAYYLQSQQKEVTIFDHGQ-QQTKAAAGIISPWFSKRNKWWYMARL 63
      I+GAGI+G++ AY+L ++G + VT+ D + QQAT AAAGI+ EW S+REN+ WY++A+
Sbjct: 6 IIGAGIIGASTAYHAKTIGAR-VTVIDRKSPQATDAAAGIVCPWLSQRNKKQWYQLAR 64

Query: 64 GADFYQQLINDLRDGFATDFYQQNGIYVLKQKBEKLRDLVLAARVKSPIGLAIK 123
      GA +Y+ LI+ L+DG + Y++ G + KL + EA R+ ++P IG++
Sbjct: 65 GARYYKDLIHQLEKDGESDTGYKKVGALSIHTDASKLDMESRAYKREDAPEIDITRL 124

Query: 124 NRKELGNDFFKLGIFDNLCLASGAARVSGAALCETLLKAS--GYPVIRQKPTLKQQG-- 178
  
```

-960-

```

      + E F L ++ SGAARV G ALC +LL A+ G VI+ +L +
Sbjct: 125 SASSTKCLFPILADGVESVHISGAARVNGRAICHELLESAARKGATVTKGNASILLFENGTF 184
Query: 179 -SGVETIAGHYF--DQVITANGAWLPDLRLPLGYQVDVVRPQKQQLLDYDVHHIISDTYTPVY 235
      +G+ F D VI+ AGAW ++L+PLG V QK Q++ +++ + ++PVV
Sbjct: 185 VTGVQTDTKQFAADAVITVINGAWANEILKPLSHFQVSPQKQIMHFEMTDADVTSKWPVY 244
Query: 236 MPEGEIDILTFPNQGGKISVGTSHENDKGY-DLEPQWQVLKKLEMQALTYLPLKREMTQKTC 294
      MP + ++ P+ G+I G +HEND G DL ++ +AL P L +A
Sbjct: 245 NPPSDQYILSPONGRIVAGATEENDAGLDLDRVITGGQHEVLKALAVAPGLADRAAVET 304
Query: 295 RVGIRAYTDEDSPFYGVQVGLKNIYTAGSLGSSGLTVGVLIGVRELQILLGEGLLTPSD 354
      RVG R +T + P G V ++ LY A+GLG+SLGT+GP +G ELA+L+LG + L S
Sbjct: 305 RVGFRPFTPGFLPVVGAIVPVQKLYAANKLGASKLITWKFPLGARIKALVLGKQTEILSP 364
Query: 355 YSP 357
      Y P
Sbjct: 365 YDP 367

```

20 An alignment of the GAS and GBS proteins is shown below.

Identities = 150/172 (87%), Positives = 161/172 (93%)

```

Query: 1 MDLNNYIASIENYPQGGITFRDISPLMADGKAYSAYVREIVQYAADKIDIMVGFPEARGF 60
MDL NYIASI++YP+ GITFRDISPLMADGKAYSIA+REI QYA DNDIDM+VGFPEARGF
25 Sbjct: 1 MDLNTYIASIKDYFKAGITFRDISPLMADGKAYSIAIRRIQAACDNDIMVGFPEARGF 60
Query: 61 IVGCFVAYALGIGFAPVRKPGKLPREVIADYEKEYGLDTLTTHADAIKPGQRVLVIDDL 120
I+GCFVA LGIGFAPVRKPGKLP+V+SADYEKEYGLDTLTTHADAIKPGQRVLVIDDL
30 Sbjct: 61 IIGCFVAVELGIGFAPVRKPGKLPD+V+SADYEKEYGLDTLTTHADAIKPGQRVLVIDDL 120
Query: 121 LATGGTVKATIEIMIEKLG+VAGCAFL+VELDGLNGRKAIBGYDTKVLNMPFG 172
LATGGTVKATIEIMIEKLG+VAGCAFL+EL+GLNGR AI YD KVLN PFG
Sbjct: 121 LATGGTVKATIEIMIEKLGIVAGCAFLIELEGLNGRHAIRNYDTKVLNQFPFG 172

```

35 SEQ ID 2638 (GBS419) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 6; MW 22.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 172 (lane 4; MW 47.5kDa).

GBS419-GST was purified as shown in Figure 219, lane 6-8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 871

A DNA sequence (GBSx0923) was identified in *S.agalactiae* <SEQ ID 2641> which encodes the amino acid sequence <SEQ ID 2642>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0847 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
50 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAAL1244 GB:D78182 ORF2 [Streptococcus mutans]
Identities = 140/225 (62%), Positives = 178/225 (78%)
55 Query: 1 MTYLSQYQSGQLTLPSALFPHFKSIFKTTADDFLVNQFFYLQNTIMLSDLTSPRIATSLDK 60
      M++L+ Y+SG L LPSAL FH+K IF ADDFLVNQFFY QNTI + D+ S+IRT++ K

```

-961-

Sbjct: 1 MSFLQHYKSGNLVLP SALLPHYKDIPSNADDFLWQFFYFQNTTMMEDIATSQIATAIGK 60

Query: 61 TVADINRSISNLT SQGLLOVKTIELNHSIRIIFDTSVPFAKLDKLFEDNQVIIDNKTS 120
TV ++NRS+SNLT SQ LLD+KTIEL+ R R++FD + KLD L ++ + +

5 Sbjct: 61 TVPEVNRVS NLT SQGLLDKMTIELDGESEVLFPDATALAKKLDLDTAADETVSSSGKT 120

Query: 121 SNRLKDLVDFERELGRLLSPFEEDLQKTLQSDQTDPDIVRAALREAVFNGKTSNNYIN 180
SN LKDLV DFERELGR+LSPFEEDLQKT+ +D+TDPD+VR+ALREAVFNGKT+NNYI

10 Sbjct: 121 SNALKDLVDFERELGRMLSPFEEDLQKTVSDDKTDPDLVRSAALREAVFNGKTNWYIQ 180

Query: 181 AILRWRRREGITTLRQVEERRKQEPDNQMKDLAISDFPNAMMLW 225
AILRWRRREG++TLRQVEER++ RE ++ +SDF +AMMLW

Sbjct: 181 AILRWRRREGISTLRQVEERRKEREQANPANVTSDPFLSAMMLW 225

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2643> which encodes the amino acid sequence <SEQ ID 2644>. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have a cleavable N-term signal seq.

20 ----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

25

The protein has homology with the following sequences in the databases:

>GP:BAAL1244 GB:D78182 ORP2 [Streptococcus mutans]
Identities = 154/228 (67%), Positives = 188/228 (81%), Gaps = 1/228 (0%)

30 Query: 1 MSFLQHYKSGNLVLP SALLPHYKDIPSNADDFLWQFFYFQNTTMRDLPASQIAHAIGK 60
MSFL+HYKSGNLV+PSALLPHYKD+P++DDFLWQFFY QNTT +D+A SQIA A+GK
Sbjct: 1 MSFLQHYKSGNLVLP SALLPHYKDIPSNADDFLWQFFYFQNTTMMEDIATSQIATAIGK 60

35 Query: 61 SVADINKIISLTMQGLDMRTIELTGHRIIFDASPLVAKLDQLFVSQTATEDIKQSE-T 119
+V ++N+ +S+L +Q LLD+KTIEL GE R++PDR+ L KLD L + T + + T
Sbjct: 61 TVPEVNRVS NLT SQGLLDKMTIELDGESEVLFPDATALAKKLDLDTAADETVSSSGKT 120

Query: 120 PNHFKRLVDFERELGRLLSPFEEDLQKTLQSDQTDPDILREALREAVFNGKTNWYIQ 179
N K LV++FERELGR LSPFEEDL+KT+ DDKTDPDL+R AL+RAVFNKTNW YIQ

40 Sbjct: 121 SNALKDLVDFERELGRMLSPFEEDLQKTVSDDKTDPDLVRSAALREAVFNGKTNWYIQ 180

Query: 180 AILRWRRREGIVNLRQVEERRRVEGEDELQVTTSEDPLSAMMLWSDS 227
AILRWRR+EGI LRQVEER++ RE + + VT+S+DPLSAMMLWSDS

45 Sbjct: 181 AILRWRRREGISTLRQVEERRKEREQANPANVTSDPFLSAMMLWSDS 228

An alignment of the GAS and GBS proteins is shown below.

Identities = 144/225 (64%), Positives = 179/225 (79%), Gaps = 1/225 (0%)

50 Query: 1 MYTLEQVQSQGLTLP SALLFPHFKSIFKTAADDFLWQFFYFQNTTINSLDTPSRIATSLDK 60
M++LE Y+SG L +PSAL FH+K +R++++DDFLWQFFYFQNTT DL PS+IA ++L K
Sbjct: 1 MSFLQHYKSGNLVLP SALLPHYKDLFKSSDDFLWQFFYFQNTTMRDLPASQIAHAIGK 60

Query: 61 TVADINRSISNLT SQGLLOVKTIELNHSIRIIFDTSVPFAKLDKLFEDNQVIIDNKTS 120
+VADIN+ IS+LT+QGLD++TIEL EIBIIFD SPV AKLD+LF ID K

55 Sbjct: 61 SVADINKIISLTMQGLDMRTIELTGHRIIFDASPLVAKLDQLFVSQTATEDIKQSE-T 119

Query: 121 SNRLKDLVDFERELGRLLSPFEEDLQKTLQSDQTDPDIVRAALREAVFNGKTSNNYIN 180
N K LV +FERELGR LSPFEEDL+KTL++D+TDPD++R AL+RAVFNKTI+W YI

60 Sbjct: 120 PNHFKRLVDFERELGRLLSPFEEDLQKTLQSDQTDPDILREALREAVFNGKTNWYIQ 179

Query: 181 AILRWRRREGITTLRQVEERRKQEPDNQMKDLAISDFPNAMMLW 225
AILRWRR+BG+ LRQ+ER++ RE + + IS+DF +AMMLW

Sbjct: 180 AILRWRRREGIVNLRQVEERRRVEGEDELQVTTSEDPLSAMMLWSDS 224

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 872

A DNA sequence (GBSx0924) was identified in *S.agalactiae* <SEQ ID 2645> which encodes the amino acid sequence <SEQ ID 2646>. Analysis of this protein sequence reveals the following:

Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.1617 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAAL1245 GB:D78182 GRF3 [Streptococcus mutans]
Identities = 134/226 (59%), Positives = 170/226 (74%)

Query: 2 DLQLSKRLQKVANYVPKGRLLDVGSDHAYLPIFLCMGYCDFAIAGEVNVGPGYSALK 61
++ LS RLQ+VA++VPKGRLLDVGSDHAYLPI+LL+ G DFA+AGE++ GPY+SA+ N
Sbjct: 7 EVLSHRLQEVASFVPKGRLLDVGSDHAYLPIYLLEQLIDPAVAGEIILKPYESAVAN 66

Query: 62 VSEHGLTSKIDVRLANGLSAFEEADNIDITITICMGGRILIADILNNDIKLQHVKTIVLQ 121
V+E GL+ +I VRLA+GL+A +D+D ITICMGGRILIADIL DKL VK L+LQ
Sbjct: 67 VNESGLSQIATVRLADGLAALNNDIDITITICMGGRILIADILAGSDRLNSVQLILQ 126

Query: 122 PNNREDDLRKWLAAANDFPIVAEDILTENDKRYEILVVKHGHMNLTAKEIRPGPFLSNNT 181
PNN EDDLR WL ANDF I AE ++ + K YEILVV+ G + L+ K+LRPGPFL +
Sbjct: 127 PNNCEDDLRLVANDFMKAEKWKVRHKYKYEILVVEKGIITLSDKDLRPGPFLRQERS 186

Query: 182 TVFKERWQNELNKLTFALNSIPNSIMREERAILDEKIQDIKEVLDES 227
++FKE+W+ EL KL AL +P K + L' Ki+ I+EVL ES
Sbjct: 187 SIFKRWKRELAKLEALTRVPAKKADNMFLSTKIEQIREVLYES 232

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2647> which encodes the amino acid sequence <SEQ ID 2648>. Analysis of this protein sequence reveals the following:

Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.0803 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 145/224 (64%), Positives = 173/224 (76%)

Query: 1 MDQLSKRLQKVANYVPKGRLLDVGSDHAYLPIFLCMGYCDFAIAGEVNVGPGYSALK 60
MD QLS RL +A YVVPKG+LLDVGSDHAYLPIFL+ AIAAGEVV GPY+SA+K
Sbjct: 1 MDSQLSNRLQCAVAYVPKGVKLLDVGSDHAYLPIFLVETINQISAIAAGEVNVGPGYSALK 60

Query: 61 NVSEHGLTSKIDVRLANGLSAFEEADNIDITITICMGGRILIADILNNDIKLQHVKTIVL 120
NV++ GL I VRLANGLSAFEEAD++ ITICMGGRILIADIL +KIQ ++ LVL
Sbjct: 61 NVITQSGLAHEIIVRLANGLSAFEEADVDITITICMGGRILIADILAGKELKQIERIVL 120

Query: 121 QPNNREDDLRKWLAAANDFPIVAEDILTENDKRYEILVVKHGHMNLTAKEIRPGPFLSNNT 180
QPNNREDDLR WL+ N P+IVAE I+ ENDK YEI+V +HG L+A ELRPGP+L
Sbjct: 121 QPNNREDDLRWLAVNAPKIVARTIMAEKNKYEILVVAEGEALSAETLRNGPYLSQEK 180

Query: 181 TVFKERWQNELNKLTFALNSIPNSIMREERAILDEKIQDIKEVL 224

-963-

Subject: 181 SVVFKEKXWOREMDKLAYALSCIPREKTOEROLLLTKIQIQIKEVI 224

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 873

A DNA sequence (GBSx0925) was identified in *S.galactiae* <SEQ ID 2649> which encodes the amino acid sequence <SEQ ID 2650>. Analysis of this protein sequence reveals the following:

```

Possible site: 54
10 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3245(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
15      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9893> which encodes amino acid sequence <SEQ ID 9894> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

20 >GP.BAAL1246 QB:D78182 ORF4 [Streptococcus mutans]
   Identifies = 187/262 (71%), Positives = 224/262 (85%)

Query: 2 MKARELDVYETYPQPGESMEGDISGLQIGSLDKRIKTINVALDVRETTVAEAEIKQVD 61
      MKA ++I YE YCPQ+LS+EDGISGLQIG+LDEIK+ ++ALDVRETTVAEAEI++VLD
25 Sbjct: 1 MKASQIKRIYKAYCPQDLS+EDGISGLQIGTLDKEIKRIMIALDVRETTVAEAEIKQVD 60

Query: 62 LIVKAIIFRFLKDLVATPNQIKYIDLLSLDVAVVSHNTIDIVGLNGWFCCLDIDQY 121
      LIVKAIIFRFLK+LV T QN IY +L+K DIAVVSHTNIDIV+GLN+DWC+LDDI+
30 Sbjct: 61 LIVKAIIFRFLNLDVATQNHVIFLDEKIDAVVSHNTIDIVGGLN+DWCFLDLDI 120

Query: 122 PDILSETSGYIGIRIGRIDIPQSYEPFPAWKIKDVGGLDSVRLVSDKNPEIQVAICGG 181
      ILS + + YGIGR+DIP S EPE A K+K + F LDSGLRVSY ++NP I R+AIICGG
35 Sbjct: 121 RRLISPSKLDIGIRIGVGDIFSLKAGKVKYFIDSLRGLVSDKNPEIADICGG 180

Query: 182 SQGSFYGEAKIARAGDVFTGDIYYHTAQEMTNGLLADPGHHIEVLVFKSLATMIEWK 241
      SQGSFY+EA+ KSA Y+TGDIYYHTAQEM+TNGLLA+DPGHHIEVLV K+A + W
40 Sbjct: 181 SQGSFYQALTKAQYVITGCTYYHTAQEM+TNGLLALDPGHHIEVLVFKRLARKPQIWS 240

Query: 242 LEKGGDSVLSLEKAPTNPFFYM 263
      ++WDI++LS+ TNPFFH+
Sbjct: 241 COENNDITLESOVNTNPFFYL 262

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2651> which encodes the amino acid sequence <SEQ ID 2652>. Analysis of this protein sequence reveals the following:

```

45 possible site: 53
   >>> Seems to have no N-terminal signal sequence

----- Final Results -----
50 bacterial cytoplasm --- Certainty=0.1804 (Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 169/262 (64%), Positives = 214/262 (81%)

Query: 2 MKARRRILIVYETYPQGLSMBGDISGLQIGSLDKBKITVMVALDVRETTVAEAIERQVGL 61
 MKA+ LID YE +CP +LSMRGD+ GLQ+GSLDK+I VM+ LD+RR+TVAEAI+ +VDL

-964-

Sbjct: 3 MKAKTLIDAYRAFCFIDLSMEGDVKGQMSLQDKIRKVMITLDIRESTVAEAIKNEVDL 62
 Query: 62 LIVKHAIPFRLKDLVATPQNKIYIDLLKSDIAVVSHTNIDIVPGLNDWFCLELDIQY 121
 +I KHAIPF+PLEDLV++PQ I +DL+K DI+VVSHTNIDIVP GLNDWFCLELDI+ 1+
 5 Sbjct: 63 IITKHAPIFKPLEDLVSSPQRDILLDLVGHIDIVSVSHTNIDIVPGLNDWFCLELEKE 122
 Query: 122 PDILSETSNYGIGIRIGDIRPQSEPPFAMKIDVFGLDSEVRLVSYDKSNPRIQVAIOGG 181
 LSET G+GIGRIC ++ Q+ E A K+K VF LD+VRL+ YDK NP I +AIOGG
 10 Sbjct: 123 ATYLSETKEGFGIGRIGITVKEQALEELASKVKRVFDLDTVRLIRYDKENPLISKIAIOGG 182
 Query: 182 SGQSPYKCAIKAGADVFTGDIYYHTAQEMITNGLAIDPGHHIEVLFPVSKIAIMEQNK 241
 SG FY++A+ KGADV+TCDIYYHTAQEM+T GL A+DPGHHIEVLFP K+ ++ NK
 Sbjct: 183 SGGEFYQDAVQKGADVFTGDIYYHTAQEMTBSLFAVDPGHHIEVLFTSKLEKLGQWK 242
 15 Query: 242 LEKGNWISVLESKAPTNPFFVM 263
 E GND+S+ SKA TNEF H+
 Sbjct: 243 BEKGNWDSIISKASTNPFFSHL 264

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 874

A DNA sequence (GBSx0926) was identified in *S. agalactiae* <SEQ ID 2653> which encodes the amino acid sequence <SEQ ID 2654>. This protein is predicted to be (). Analysis of this protein sequence reveals the following:

25 Possible site: 41
 >>> Seems to have a cleavable N-term signal seq.
 ----- Final Results -----
 30 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP: CAB15253 GB: Z99120 similar to opine catabolism [Bacillus subtilis]
 Identities = 148/368 (40%), Positives = 211/368 (57%), Gaps = 13/368 (3%)
 Query: 1 NKKIAIAGAGAVGATLAYLSKEKDVIQVTFDYG+V- QGATGAAAGIISPFWSEKRRNIGWY 59
 NK I+GAG +GA+ AY+L+K +VTV D QGAT AAGI+ FW S+RNN+ WY
 Sbjct: 1 MKSYIIIVAGILGASTAYHLAKT- GARVTVIDRKEFGQATDAAGIVCPNLQRRNQWY 59
 40 Query: 60 RMARLGADFYSLAVTDLQKQGFETKFTQQTGVFLKKDESQLESIFALADKRRLESPLIG 119
 ++A+ GA +Y L+ L+KDG Y+ G + D S+L+ + A KRR ++P IG
 Sbjct: 60 QLAKGAGRYVYKDLHLQLEKDGESDTQYKRVGAISHTDASKLDMREYAYKRRDAPEIG 119
 45 Query: 120 DLQILNKSEANTHPPEL-DGYEQLLYASGGARVSGADLFRILLEAS---GVNVIKDEVHF 175
 D+ L+ SE FP L DGYE Y+ SG ARV G L R LL A+ G VIK
 Sbjct: 120 DITRLASSTFKLPFLILDGYES-VHISGAARVNGRALCRSLSSAEKRGATVKEGNASL 178
 Query: 176 -----TITDNGFRVGHIDFDELVLASGAMLAKILDEHNIVQDVDPQKGLRQDYFNFINT 230
 T+T + D +++ +GAW +LL V QK Q+ + ++ +T
 50 Sbjct: 179 LFENGTVTGVTDTKQEAADAVITAGAWANELLKPLGLHFGQVPSQKQAMHFFMTADT 238
 Query: 231 GKYPVUMPGEGLDIIPFDNGKVSVAASHNEMAF-DLNIDFKVLDFKPERQAGFYFQGLKK 289
 G +PVUMP + I+ FDNG++ GA+HEND DL + + +A+ P L
 55 Sbjct: 239 GSWPVMPPSPDQYILSFDNKRIVAGATHENDAGLDLDRVTAGQHQEVLKSLAVAGDEL 298
 Query: 290 ADTTSERVGIRAYTSDPSFPFGVPVPMGAGYASGLSTGLTGVPLIGYELQGLLNKEN 349
 A RVG R +T F F G VP ++G YAA+GLG++GLT+GP +G EL +L+L K+
 Sbjct: 299 AAARVETRGFRFTTGPGLFVVGAVPVGGLYAAAGLGASGLNGPFLGLAKLVLGQY 358
 60 Query: 350 QLNLEDYD 357

-965-

+LaL YD
 Sbjct: 359 ELDLSPD 366

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2655> which encodes the amino acid sequence <SEQ ID 2656>. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 211/360 (58%), Positives = 262/360 (72%)

Query: 3 KIAIIGAGAVGATLAIYLSKEKDIQTVFVYGVQATKAAAGIISWFSKRRNKMYRMA 62
 KIAIIGAG VG+I AYFL + +VT+FD+G GQATKAAAGIISWFSKRRNK MYRMA
 Sbjct: 2 KIAIIGAGIVGSTAAIYLCQSQGKEVITFDHGGQATKAAAGIISWFSKRRNKMYRMA 61

Query: 63 RLGADFYSLVTDLQKDGFEFKFYQGTGVFLKKEDSQLESFALADKKRRLSPGIDLQ 122
 RLGADFY +L+ DL++DGF T FKQ G+++LKK E +L L+ LA R++ESP+IG+L
 Sbjct: 62 RLGADFYQLINDLKEDGFATDFYQGNIVYLVKKQBEKRLDLYELALARKVESP+IGELA 121

Query: 123 IANKSEANTHFPPELDGYBQLLYASGGARVEGADLIRILLEASGVNVIKDEVHFTITDNGF 192
 I N+ E F L G++ LYASG ARVEGA L LL+ASG VI+ +V +G+
 Sbjct: 122 IKNRKEIGNDFKGLIGFNCLYASGAARVEGAALCETLLKASSYPTVHQKVTYKQSSGY 181

Query: 183 RVQVIDFQKVLASGAWLAKIIDEHNYQVDVRPQKGQLRDYFNSINTKYPVVMPEGEL 242
 + G FD++LA+GAWL +L YQVDVRPQKGQ DY +I + YPVVMPEGE+
 Sbjct: 182 ETAGRYFDQVILANGAWLPDLIRPLGYQVDVRPQKGQLLDYDVHHSIDTYFPVVMPEGEI 241

Query: 243 DIIPFNKRVSVGASHENEMAFDLNIDFKVLDFKEBQAGVFPOLAKADTTSERVGIKAY 302
 D+IFF+ GK+SVG SHEND +DL D++VL K E QN+ Y F LK+A + RVGIKAY
 Sbjct: 242 DLIPFNQKISVGTSHENDKGYDLRFQWCVLKKLEWQALTYLPLLEKATQKTCRVGIKAY 301

Query: 303 TSDPSPPFGQVPPOMEGAYASGLSGTGLTVGLIGYELQQLINKENQINLEDDVDITKYV 362
 TSD+SPF+G V ++ Y ASGLSG+GLTVGLIGYEL QL+L E L DY Y+
 Sbjct: 302 TSDYSPPYQVSGLQNLTYASGLSGSLTVGLIGYELQQLLHGEGLLTSDYSPEPYL 361

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8679> and protein <SEQ ID 8680> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 2
 MOG: Discrim Score: 4.44
 CvH: Signal Score (-7.5): 0.81
 Possible site: 41
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 0 value: 7.32 threshold: 0.0
 PERIPHERAL Likelihood = 7.32 153
 modified ALOM score: -1.96
 +++ Reasoning Step: 3

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

-967-

Query: 1 MKGIILAGSGSTRLYPLTRAASKQLMPYDKPMIYYPLSLVLMAGIKRILIIISTPODLPR 60
 MKGIILAGSGSTRLYPLTRAASKQLMPYDKPMIYYPLSLVLMAGIKRILIIISTPODLPR
 Sbjct: 1 MKGIILAGSGSTRLYPLTRAASKQLMPYDKPMIYYPLSLVLMAGIKRILIIISTPODLPR 60

5 Query: 61 FEMILGDGSEKLGISLSYARQSPDGLAQAFIIGREFIGDDHVALVIGDNIYHGGPGLSAML 120
 P+D+L DGE GI LSYARQSPDGLAQAFIIGREFIGDDHVALVIGDNIYHGGPGLSAML
 Sbjct: 61 FEMILGDGSEKLGISLSYARQSPDGLAQAFIIGREFIGDDHVALVIGDNIYHGGPGLSAML 120

10 Query: 121 QRAAKKSGATVFGYQVKDPERFGVVFEDIMNAISIEEKPAQPKSNYAVTGLYFYDNDV 180
 Q+AA KE GATVFGYQVKDPERFGVVFEDIMNAISIEEK P+SNYAVTGLYFYDNDV
 Sbjct: 121 QRAAKKSGATVFGYQVKDPERFGVVFEDIMNAISIEEKPEYPRSNYAVTGLYFYDNDV 180

15 Query: 181 VEIAKNIKPSRGELEITTVNKAYLDRGDSVLMGRGFAMLDTGTHESLLEAAQYIETV 240
 VEIAK IKPS RGELEITTVNKAYL+RGDSVLMGRGFAMLDTGTHESLLEAAQYIETV
 Sbjct: 181 VEIAKNIKPSRGELEITTVNKAYLDRGDSVLMGRGFAMLDTGTHESLLEAAQYIETV 240

20 Query: 241 QRMQNVQVANLEEIAYRMGYISREDVLELAQSLKNEYGYQLRLRIGEA 289
 QRMQNVQVANLEEI+YRMGYI+RE VLELAQSLKNEYGY+QLRLRIGEA
 Sbjct: 241 QRMQNVQVANLEEISYRMGYISREDVLELAQSLKNEYGYQLRLRIGEA 289

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2659> which encodes the amino acid sequence <SEQ ID 2660>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1585 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30 RGD motif: 207-209

The protein has homology with the following sequences in the databases:

>GP:AAAC69538 GB:AP057294 Cpe23f0 [Streptococcus pneumoniae]
 Identities = 263/289 (91%), Positives = 276/289 (95%)

35 Query: 1 MKGIILAGSGSTRLYPLTRAASKQLMPYDKPMIYYPLSLVLMAGIKRILIIISTPODLPR 60
 MKGIILAGSGSTRLYPLTRAASKQLMPYDKPMIYYPLSLVLMAGIKRILIIISTPODLPR
 Sbjct: 1 MKGIILAGSGSTRLYPLTRAASKQLMPYDKPMIYYPLSLVLMAGIKRILIIISTPODLPR 60

40 Query: 61 FEMILGDGSEKLGISLSYKBPSPDGLAQAFIIGREFIGDDHVALVIGDNIYHGGPGLSAML 120
 F+ELL DGEFQGI LSY EBPSPDGLAQAFIIGREFIGDDHVALVIGDNIYHGGPGLSAML
 Sbjct: 61 FEMILGDGSEKLGISLSYKBPSPDGLAQAFIIGREFIGDDHVALVIGDNIYHGGPGLSAML 120

45 Query: 121 QRAAKKSGATVFGYQVKDPERFGVVFEDIMNAISIEEKPEVPRSNYAVTGLYFYDNDV 180
 QKAA KEGATVFGY VKDPERFGVVFEDIMNAISIEEK P+S+AVTGLYFYDNDV
 Sbjct: 121 QRAAKKSGATVFGYQVKDPERFGVVFEDIMNAISIEEKPEYPRSNYAVTGLYFYDNDV 180

50 Query: 181 VEIAKNIKPSRGELEITTVNKAYLDRGDSVLMGRGFAMLDTGTHESLLEAAQYIETV 240
 VEIAK+IKPS RGELEITTVNKAYL+RGDSVLMGRGFAMLDTGTHESLLEAAQYIETV
 Sbjct: 181 VEIAKNIKPSRGELEITTVNKAYLDRGDSVLMGRGFAMLDTGTHESLLEAAQYIETV 240

55 Query: 241 QRLQNAVANLEERIAIRMGYISKEDVHKLQSLKNEYGYQLRLRIGEA 289
 QR+QN QVANLEERIAIRMGYIS+EDV LQSLKNEYGYQLRLRIGEA
 Sbjct: 241 QRMQNVQVANLEEIAYRMGYISREDVLELAQSLKNEYGYQLRLRIGEA 289

An alignment of the GAS and GBS proteins is shown below.

Identities = 257/289 (88%), Positives = 274/289 (93%)

60 Query: 1 MKGIILAGSGSTRLYPLTRAASKQLMPYDKPMIYYPLSLVLMAGIKRILIIISTPODLPR 60
 MKGIILAGSGSTRLYPLTRAASKQLMPYDKPMIYYPLSLVLMAGIKRILIIISTPODLPR
 Sbjct: 1 MKGIILAGSGSTRLYPLTRAASKQLMPYDKPMIYYPLSLVLMAGIKRILIIISTPODLPR 60

Query: 61 FEMILGDGSEKLGISLSYARQSPDGLAQAFIIGREFIGDDHVALVIGDNIYHGGPGLSAML 120

-968-

```

PE++LGDGSE GISISY EQSPDGLAQAPLIGR+FIGDD VAL+LGNIIYHG GI+ M1
Sbjct: 61 PEELLGDGSEPGISLGYKQPSDGLAQAPLIGRFIGDDVALILGDNIIYHNGILTRML 120

Query: 121 QRAASKRSGATVPGYQVKDPKRGVVEPITDMNAISIEKKPAQPSKNYAVTGLVFFDNV 180
5 Q+A+KE GATVPGYQVKDPKRGVVEPD +MNAISIEKK PKS++AVTGLVFFDNV
Sbjct: 121 QKAAAKERKATVPGYQVKDPKRGVVEPDENMNAISIEKKPEVVKSHFAVTLVFFDNV 180

Query: 181 VETAKNIKPSRGELEITDVKAYLERGDLSELMRGGFANLDTGTHESLLEAAQYIETV 240
VETAKNIKPS RGELEITDVKAYL+RGDLSELMRGGFANLDTGTHESLLEAAQYIETV
10 Sbjct: 181 VETAKNIKPSRGELEITDVKAYLERGDLSELMRGGFANLDTGTHESLLEAAQYIETV 240

Query: 241 QR+QN QVANLEETAYRMGYITREQVLELAQPLKNEYQYLLRLIGEA 289
QR+QN QVANLEETAYRMGYI++E V +LAQ LKNEYQYLLRLIGEA
15 Sbjct: 241 QRLQNAQVANLEETAYRMGYISKEDVHKLAQSLKNEYQYLLRLIGEA 289

```

There is also homology to SEQ ID 858.

SEQ ID 2658 (GBS296) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 43 (lane 5; MW 35.4kDa).

GBS296-His was purified as shown in Figure 203, lane 7.

- 20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 876

A DNA sequence (GBSx0929) was identified in *S. agalactiae* <SEQ ID 2661> which encodes the amino acid sequence <SEQ ID 2662>. Analysis of this protein sequence reveals the following:

```

25 Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
30 bacterial cytoplasm --- Certainty=0.2635(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 877

A DNA sequence (GBSx0930) was identified in *S. agalactiae* <SEQ ID 2663> which encodes the amino acid sequence <SEQ ID 2664>. This protein is predicted to be unnamed protein product. Analysis of this

```

40 protein sequence reveals the following:

Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
45 bacterial cytoplasm --- Certainty=0.1868(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2665> which encodes the amino acid sequence <SEQ ID 2666>. Analysis of this protein sequence reveals the following:

50

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2818 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

RGD motif: 29-31

The protein has homology with the following sequences in the databases:

>GP:AA069539 GB:AF057294 Cps23FP [Streptococcus pneumoniae]
 Identities = 168/197 (85%), Positives = 183/197 (92%)

Query: 1 MTETFFDKPLACREIKETPGLLEFDIPVRGDNRGWFKENFQKEKMLP+IGFPERFFBEGKL 60
 MT+ FF K LA R+++ IPG+LEFDIPV GDNRGWFKENFQKEKMLP+GFPE FF BSKL
 Sbjct: 1 MTDNFGKTLAARKVSAIPGMLEFDIPVHGDNRGWFKENFQKEKMLPLGFPESFFABEGKL 60

Query: 61 QNNVFSRQHVLRGLHAEPWDKYSVADGGKVLGAWVDLREGSTFGNVTQVIDASKGMF 120
 QNNVFSR++YLRGLHAEPWDKYSVAD GKVLG+YVDLREGSTFGN YQTVIDASKG+F
 Sbjct: 61 QNNVFSRKNVLRGLHAEPWDKYSVADGGKVLGSMVDLREGSTFGNTYQTVIDASKGIF 120

Query: 121 VPRGVANGFQVLSSETVSYSYLNDYWDALDLKPKYAFVNTADPSLGIWTMENLAAAEVSEAD 180
 VPRGVANGFQVLS+TVSYSYLNDYWDAL+LKPKYAFVNTADPSLGI WEN+AA AEVSEAD
 Sbjct: 121 VPRGVANGFQVLSDTVSYSYLNDYWDALDLKPKYAFVNTADPSLGIWENLAAAEVSEAD 180

Query: 181 KNHPLLSDVKPLKPKDL 197
 K+HPLL DVKPLK +DL
 Sbjct: 181 KNHPLLDVVKPLKPKDL 197

An alignment of the GAS and GBS proteins is shown below.

Identities = 157/197 (79%), Positives = 180/197 (90%)

Query: 1 MTEQFFDKELTCRPIAIPGLLEFDIPVRGDNRGWFKENFQKEKMIPLGFPESFFBEGKL 60
 MTE FFDK L CR I+ IPGLLEFDIPVRGDNRGWFKENFQKEKMI+P+GFPS FFE KL
 Sbjct: 1 MTETFFDKPLACREIKETPGLLEFDIPVRGDNRGWFKENFQKEKMLP+IGFPERFFBEGKL 60

Query: 61 QNNISFNKNITLRLGLHAEPWDKYSVADGGKVLGAWVDLREGSTFGNVTQVIDASKGIF 120
 QNN+SF+++ LRGLHAEPWDKY+S+AD+G+V+G WVDLREG++FGNVTQ+IDASKG+F
 Sbjct: 61 QNNVFSRQHVLRGLHAEPWDKYSVADGGKVLGAWVDLREGSTFGNVTQVIDASKGMF 120

Query: 121 VPRGVANGFQVLSOKAAYTLYNDYWDALDLKPKYAFVNTADPNLGIQWENLSEAEVSEAD 180
 VPRGVANGFQVLS+ +Y+YLNDYWDAL+LKPKYAFVNTAD+LGI WENL AEVSEAD
 Sbjct: 121 VPRGVANGFQVLSSTVSYSYLNDYWDALDLKPKYAFVNTADPSLGIWTMENLAAAEVSEAD 180

Query: 181 KNHPLLSDVKPLKPKDL 197
 KNHPLL DVKPLK +DL
 Sbjct: 181 KNHPLLSDVKPLKPKDL 197

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 878

A DNA sequence (GBSx0931) was identified in *S.galactiae* <SEQ ID 2667> which encodes the amino acid sequence <SEQ ID 2668>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3019 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 879

A DNA sequence (GBSx0932) was identified in *S.agalactiae* <SEQ ID 2669> which encodes the amino acid sequence <SEQ ID 2670>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 880

A DNA sequence (GBSx0933) was identified in *S.agalactiae* <SEQ ID 2671> which encodes the amino acid sequence <SEQ ID 2672>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0957 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9367> which encodes amino acid sequence <SEQ ID 9368> was also identified.

The protein is similar to the dTDP-glucose-4,6-dehydratase from *S.mutans*:

```
>GP:BAAL1249 GB:D78182 dTDP-glucose-4,6-dehydratase [Streptococcus mutans]
Identities = 290/310 (93%), Positives = 304/310 (97%)

Query: 1 MTYAGNKRANIEAILGDRVELVVGDIADAEILVDKLAAKADAIVHYAARSHNDNSLNDPSPF 60
      +TYAGN AN+E ILGDRVELVVGDIAD+ELVDKLAAKADAIVHYAARSHNDNSL DPSPF
Sbjct: 39 LTYAGNHNANLEILGDRVELVVGDIADSEILVDKLAAKADAIVHYAARSHNDNSLNDPSPF 98

Query: 61 IHNFIQITITLLEAARKYDIRFHHVSTDEVYGDLPREDLPCHRGPGGKFTAEITKYNP 120
      I+TNF+GTYITLLEAARKYDIRFHHVSTDEVYGDLPREDLPCHRGPGKFTAEITKYNP
Sbjct: 99 IYINFGVITITLLEAARKYDIRFHHVSTDEVYGDLPREDLPCHRGPGKFTAEITKYNP 158

Query: 121 SPYSTKAASDLIVKAWVRSPGVKATISNCNNYGPYQHIEKFTIPROITNIIAGIKPKLY 180
      SPYSTKAASDLIVKAWVRSPGVKATISNCNNYGPYQHIEKFTIPROITNIIAGIKPKLY
Sbjct: 159 SPYSTKAASDLIVKAWVRSPGVKATISNCNNYGPYQHIEKFTIPROITNIIAGIKPKLY 218

Query: 181 GEGKNVRDWHITNDHSTGVWAILTKGRIGETTYLIGADGKNNKEVLELILEKMGQPIQAY 240
```

-971-

```

      GEGQVRDWHITNDHSTGVWAILTKGRIGETYLIGADGEKNNKEVLELILEKM QPKDAY
Sbjct: 219 GEGQVRDWHITNDHSTGVWAILTKGRIGETYLIGADGEKNNKEVLELILEKM QPKDAY 278

      Query: 241 DHVTRAGHDLRYAIDSTKLREELGWEPQFINPSEGLEETINWYTENQDWWKAEKRAVEA 300
      DHVTRAGHDLRYAIDSTKLREELGW+PQFINP EGLE+TI WYTE+ +DWWKAEKRAVEA
Sbjct: 279 DHVTRAGHDLRYAIDSTKLREELGWKPQFINPSEGLEETIKWYTEHSDWWKAEKRAVEA 338

      Query: 301 NYAKTQEVIN 310
      NYAKTQ++N
Sbjct: 339 NYAKTQKIIN 348

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2673> which encodes the amino acid sequence <SEQ ID 2674>. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1150 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 300/309 (97%), Positives = 303/309 (97%)

      Query: 1  NTYAGNANIEAILGDRVELVVGDIADAELVDKLAAKADAIVHYAARSHDINSLEDPSFF 60
      Sbjct: 37  LTYAGNANIEAILGDRVELVVGDIADAELVDKLAAKTDAIVHYAARSHDINSLEDPSFF 96

      Query: 61  IHTNFIGTYTLEAARKYDIRFHVSTDEVYGDPLREDLPNGEGGPKFTASTKYNPS 120
      Sbjct: 97  IHTNFIGTYTLEAARKYDIRFHVSTDEVYGDPLREDLPQGGGPKFTASTKYNPS 156

      Query: 123  SPYSSTKAASDLIVKAWVRSPGVKATISNCNNYGPYQHIEKFIPIQITNLAGIKPKLY 180
      Sbjct: 157  SPYSSTKAASDLIVKAWVRSPGVKATISNCNNYGPYQHIEKFIPIQITNLAGIKPKLY 216

      Query: 181  GEGQVRDWHITNDHSTGVWAILTKGRIGETYLIGADGEKNNKEVLELILEKM QPKDAY 240
      Sbjct: 217  GEGQVRDWHITNDHSTGVWAILTKGRIGETYLIGADGEKNNKEVLELILEKM QPKDAY 276

      Query: 241  DHVTRAGHDLRYAIDSTKLREELGWEPQFINPSEGLEETINWYTENQDWWKAEKRAVEA 300
      Sbjct: 277  DHVTRAGHDLRYAIDSTKLREELGWEPQFINPSEGLEETIKWYTEHSDWWKAEKRAVEA 336

      Query: 301  NYAKTQEVIN 309
      YAKTQEV
Sbjct: 337  KYAKTQEVIN 345

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 881

A DNA sequence (GBSx0935) was identified in *S.agalactiae* <SEQ ID 2675> which encodes the amino acid sequence <SEQ ID 2676>. Analysis of this protein sequence reveals the following:

```

Possible site: 36
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 882

A DNA sequence (GBSx0936) was identified in *S.agalactiae* <SEQ ID 2677> which encodes the amino acid sequence <SEQ ID 2678>. Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL      Likelihood =-15.55      Transmembrane      13 - 29 ( 3 - 40)

----- Final Results -----
      bacterial membrane --- Certainty=0.7220(Affirmative) < succ>
15      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 883

A DNA sequence (GBSx0937) was identified in *S.agalactiae* <SEQ ID 2679> which encodes the amino acid sequence <SEQ ID 2680>. Analysis of this protein sequence reveals the following:

```

25      Possible site: 15
      >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2882(Affirmative) < succ>
30      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 884

A DNA sequence (GBSx0938) was identified in *S.agalactiae* <SEQ ID 2681> which encodes the amino acid sequence <SEQ ID 2682>. This protein is predicted to be hyaluronate lyase. Analysis of this protein sequence reveals the following:

```

45      Possible site: 30
      >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>

```

-973-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2683> which encodes the amino acid sequence <SEQ ID 2684>. Analysis of this protein sequence reveals the following:

Possible site: 46
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

- 10 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9099> which encodes the amino acid sequence <SEQ ID 9100>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 23
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

- 20 bacterial outside --- Certainty= 0.300 (Affirmative) < succ>
bacterial membrane --- Certainty= 0.000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty= 0.000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- 25 Identities = 359/771 (46%), Positives = 492/771 (63%), Gaps = 50/771 (6%)
- Query: 307 ENAT--GSTTKISKDKSKLLKEVPLSVTASTEDNFNFKLLDKXNDVTIGNHVYIYNISIM 364
EN T + T+ +D K+++ +D+T+LLD+WN + GN YD + +M
Sbjct: 65 ENNTYFQYQLTTTDSKKVQVE-----QQNDYYTELDDQNSIAGNDADKNTIPM 117
- 30 Query: 365 QKLNQKLDENNAKNIIEAIL-----DSNRTFLWKLDNLNLSAQIATATYRRLDIAKQIT 419
+ K E +A+NI IK NRT+LW+ + + SA +T TYR +E +AKQIT
Sbjct: 118 VTFENKA-EKDAQNI--IKSYQGFPHENTYLWEHAKDYSASANTIKTYRNIEKIAKQIT 174
- 35 Query: 420 NPHSTIYKNEKAIRTVKESLAWLHQNFYNNKDI-----EGSANMWDFEIGVPRISITGT 473
NP S Y+ KAI VK+ +A++++ YN++++ E NWW +EIG PR+I T
Sbjct: 175 NFESCYQDSKIALIVKDMAFPIYHAYNLDRHQTTGKNGKNNMWVEIGTFRAINST 234
- 40 Query: 474 LALMNYPTDAEIKTYTDPISHFVFDAGFFRKTIVN--PPKALQGNLVDMGRVKIEGGL 531
L+LNY YFT EI YT PIE FVFD FR N PF+A GNL+DMGRVK+I G+L
Sbjct: 235 LSLMYPYFQREILKYTAPEIKFVDPDTRFVRRAANPSPFENSQNLIDMGRVKILGISL 294
- Query: 532 RIGNTIIEKTSISLKNLFTTATYAEGFYADGSYIDT-----NVAYTGAIGNVL 580
RKD+ I T +++ +FT + GFY DGS IDH +AYTGAIGNVL
Sbjct: 295 RKDDLEISDTIKAEIKVFTLVDEGNFGFYQDGLSDHIVTINAQSPLYKGIAYTGAIGNVL 354
- 45 Query: 581 IDGLTOLLPIIQETDYKISNQELDMVYKWNQSFLEPLIVKGLMNSGRSISREAASSH 640
IDGL+Q+PIIQ+T I ++ +Y WIN SF P+IV+GR+MDM+GRSISR A SH
Sbjct: 355 IDGLSQLPPIQKTKSPIKADKQATYIHWINSFFPIIVRGEMMDTGRSIRFNQAQSH 414
- 50 Query: 641 AAAVEVLGRFLRLANMNEERNLDKSTIKTILTS-NKFINVFNILKSYSDIANMKNLIN 699
A +E LR LR+A+MS E L LK+ IKT++T N FYNV++NLK+Y DI M +Li+
Sbjct: 415 VAGIEFLAILRLADWSEEPHIALKTRIKTLVTQGNAPFVNYDNLKTYHDILKKEGIS 474
- 55 Query: 700 DSTVATKPLKSNLSTFNSMDRLAYYNAEKDFGFALSLHSKPTLMYKGMNDENTRGWYTD 759
D++V + L S +++FNSMD+LA YN + DF F LS+ S RT NYE MN+EN GW+T D
Sbjct: 475 DTSVPVQKLDYSVAFNSMDKALYNNKHDFAGLSNFSNRTQRYEAMNNENHGWPTSD 534
- Query: 760 GMFYLNDSQSHYSNHPWPTFVNDYQNAQTTEKDAKREDITKDFMSKSKDAKKTQVGT 819
GMFYLN+D HYS ++W TVNDY++ GTPE + K + T + + K ++ G +TG
Sbjct: 535 GMFYLYNDLGHYSYENWATVNDYKLGTTETETEQKLEGTPE---NIKINYQQQVG+MG 589
- Query: 820 ASD--FVGSVKINDHFAAAMDPTNMDRTLTAKGKGVILANDKIVFLGSIKRNINIGIRVS 877

-974-

SD FV S KLM+ ALAAM FTMW++LT KGW IL +KI+P+GSNIKN +
 Sbjct: 590 LDDA FVASKKILNTSALAAMFTNNKSLITLRKQKFLONKIIPVGSNIKNQSS -HKAY 648

Query: 878 TTIDQRKDDSKFTPTTYVNGKTVDLQASSQQPTDTRKSVFLSKKGRNIGYFFKNSTI 937
 TTI+QRK++ K PY +YVN + VDL FT+TKS+FLS +P +NIGY FFK +T+
 Sbjct: 649 TTIEQRKKNQKYPYCSYVNRQPVLDLN -QLVDFTNTRKSIFLESDDPAQNIGYFFKFTTL 707

Query: 938 DIERKESQIGTWNINSINTSKNTSI---VGNFITTSQKHNNKGDSDYDZMMVFNIDRTSPDK 994
 I + QTG W +I K+ VGN FITT Q H GD Y VM+FN+ R P+
 Sbjct: 708 SISKALQTGKNQIKADDSPEAKRVENTFITTMQNHQTQDDRYYMMGLPNTRQEPET 767

Query: 995 LANSKEVELLNSKQQVIYDRNSQTWAVIKHDKQESLINNQFTNNKAGLY 1045
 + +++LLEN+ K +YD +SQ VI + + + +N ++ G Y
 Sbjct: 768 YISKLDIDLENNDKLAAYDHDSCQMHVHYGKATMPSNH -NLSHQCFY 817

SEQ ID 2682 (GBS89) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 6 (lane 3; MW 118kDa).

The His-fusion protein was purified as shown in Figure 190, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 885

A DNA sequence (GBS0939) was identified in *Sagalactiae* <SEQ ID 2685> which encodes the amino acid sequence <SEQ ID 2686>. This protein is predicted to be mutator mutt protein. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3781(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAAL1250 GB:D78182 MutX [Streptococcus mutans]
 Identities = 132/160 (82%), Positives = 146/160 (90%), Gaps = 1/160 (0%)

Query: 1 MTKLATICYIDNGKELLRLHNNKENDVHEGKWL SVGGKLEAGETPDECAREILBETHL 60
 K LATICYIDNG+ELL+HRNKK NDVHEGKWL SVGGKLE GE+PDECA+REI BETHL
 Sbjct: 1 MTKLATICYIDNGKELLRLHNNKKNPDVHEGKWL SVGGKLEKGSPEDECAREIFETHL 60

Query: 61 TVKQDPFKVITTFPEPTFGHDWYTVVFKVIDYEGELISDDBSRGTLWVPYDVLKPT 120
 VK+MDPKG+ITFP+FTPGHDWYTVVFKV D+EG LI SD +SRGTLWVPY+QVL+KPT
 Sbjct: 61 IVKQDPFKGLITTFPDTFGHDWYTVVFKVRDFEGLISDKDSRGTLWVPYQVLTKPT 120

Query: 121 WQSDYRIFKNILEDVFFPSAKFVYDEHONLIEKTVNFYEK 160
 W+GDYRIFKNILED PPSAKFVY E Q L+K V FYEK
 Sbjct: 121 WEGDYRIFKNILEDAPPSAKFVYQD-QKLVDKHNIFYEK 159

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2687> which encodes the amino acid sequence <SEQ ID 2688>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3399(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-975-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 131/158 (82%), Positives = 146/158 (91%)

5 Query: 1 MTKLATICYIDNGKELLLHRRNKENDVHGGKWSVGGKLEAGETPDCAKREILEETHL 60
 MT+LATICYIDNG LLLHRRNKENDVH+GKWSVGGKLEAGETPDCA+REILEETHL
 Sbjct: 1 MTKLATICYIDNGSLLHRRNKENDVHGGKWSVGGKLEAGETPDCAKREILEETHL 60

10 Query: 61 TVKRMDFGQVITFPFETPGHDWYTYVFKVIDYEGELISDOESRGTLEWVPYDQVLSKPT 120
 TV +M FKG+ITFPETPGHDWYTYVFKVT +BG+LISD+BSREGTLEWVPYDQVL KPT
 Sbjct: 61 TVTEGAFGKITFPFETPGHDWYTYVFKVTGFBGLISDEBSREGTLEWVPYDQVLEKPT 120

15 Query: 121 WQGDYBIFKWLIEDVPPFSAKFVYDEHQNLIEKTVNPFY 158
 W+GEY+IFKWLIED FFSAKF YD++ L++K+V FY
 Sbjct: 121 WEGLYDIFKWLIEDRSFSSAKFTYDQNNQNDKSVTFY 158

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 886

A DNA sequence (GBSx0940) was identified in *S. agalactiae* <SEQ ID 2689> which encodes the amino acid sequence <SEQ ID 2690>. This protein is predicted to be MutT/nudix family protein. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

 bacterial cytoplasm --- Certainty=0.1901 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAFL1917 GB:AE002059 MutT/nudix family protein [Deinococcus radiodurans]
 Identities = 40/135 (29%), Positives = 62/135 (45%), Gaps = 3/135 (2%)

35 Query: 22 FGVRSALIIENQKILLIYAPHLDKTY-LPGALQVGDSDSNKAVAREVLREIGLHSCWGD 80
 F R + + + + +L L + + + LPGA+Q GR S A RE RE GL + V
 Sbjct: 33 PQTRATLICVQDNRLLTOWDERFPDPFALPGGAVQTGESSAAAAQREWHMETGLRADVTR 82

40 Query: 81 LAYIIENQPNIKRHHYHSVEFLYFVNLLQAPRESIKSGTHKRHFVMLPIKELTKIDCNPN 140
 A +E F+ + + H F + V L G+ P + + + H F NL + L P
 Sbjct: 93 CA-TLIERFFHNGCRBRHEFGFFRVELTGELPATVLDNPHV-PPRMLAVDALDQHTLYPR 150

45 Query: 141 FLAQGLIENFGHVH 155
 + Q L L G + H
 Sbjct: 151 CVPQLRLFPAGEIGH 165

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2691> which encodes the amino acid sequence <SEQ ID 2692>. Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

 bacterial cytoplasm --- Certainty=0.3832 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-976-

Identities = 33/80 (41%), Positives = 50/80 (62%), Gaps = 1/80 (1%)

Query: 29 LIENQKLLIYAPHLDKYYLPGLALQVGEDSNKAVAREVLSEIGLASQVGLAYLIENQ 88

LI+ N K L DAYY GQ VGR ++ V R LSE+G+ ++V LA+++EN

5 Sbjct: 1 LIVRNGKNFILTRDAD-DQYTTIGGTSLVGKTHETVIRETLEEVGIRAKVQLAFVNEH 59

Query: 89 FNIKHHYHSVEFLYPVNL 108

F+I +H++EF Y V+ L

10 Sbjct: 60 FDIDVPMHINIFHYLVSP 79

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 87

A DNA sequence (GBSx0941) was identified in *S. agalactiae* <SEQ ID 2693> which encodes the amino acid sequence <SEQ ID 2694>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -12.95	Transmembrane	24 - 40 (17 - 48)
INTEGRAL	Likelihood = -11.09	Transmembrane	88 - 104 (82 - 112)
INTEGRAL	Likelihood = -9.39	Transmembrane	294 - 310 (276 - 315)
INTEGRAL	Likelihood = -8.07	Transmembrane	242 - 258 (236 - 262)
INTEGRAL	Likelihood = -7.86	Transmembrane	50 - 66 (43 - 74)
INTEGRAL	Likelihood = -3.13	Transmembrane	337 - 353 (332 - 355)
INTEGRAL	Likelihood = -2.23	Transmembrane	185 - 201 (182 - 202)
INTEGRAL	Likelihood = -1.38	Transmembrane	269 - 285 (267 - 285)

----- Final Results -----

bacterial membrane	---	Certainty=0.6180 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2695> which encodes the amino acid sequence <SEQ ID 2696>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -9.71	Transmembrane	88 - 104 (85 - 112)
INTEGRAL	Likelihood = -9.29	Transmembrane	24 - 40 (21 - 72)
INTEGRAL	Likelihood = -8.92	Transmembrane	47 - 63 (41 - 72)
INTEGRAL	Likelihood = -7.59	Transmembrane	243 - 259 (237 - 266)
INTEGRAL	Likelihood = -6.10	Transmembrane	181 - 197 (178 - 203)
INTEGRAL	Likelihood = -5.47	Transmembrane	278 - 294 (273 - 310)
INTEGRAL	Likelihood = -3.88	Transmembrane	338 - 354 (331 - 368)
INTEGRAL	Likelihood = -1.59	Transmembrane	297 - 313 (297 - 314)

----- Final Results -----

bacterial membrane	---	Certainty=0.4885 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

The protein has homology with the following sequences in the databases:

>GF:AAD00285 GB:U78604 putative membrane protein [Streptococcus mutans]

Identities = 244/382 (63%), Positives = 310/382 (80%), Gaps = 3/382 (0%)

Query: 12 SLFYKFLNNQATMALVITLLAFLITFVFKISFLFMPVISFFAVIMPLVISTILYLYL 71

S F+KWEFL+N+ I++ LL FL I VFKIS +F P++SF AVIMPLVIS +LYYL

Sbjct: 17 SWFFKFLNNKVTVLVLLVFLDLVFLTKISSFKPLSLFVIMPLVISTILYLYL 76

Query: 72 KPLVLINHLGNRTSIFIVGLITLLFVWISGFVPMVQQLTSFIEDLPKYGVKNE 131

-977-

KP+VD I G +R -I IVF +I L VW I+ F PM+ QLTSTFI+ LP YV V+
 Sbjct: 77 KPVIDFIEIRGTSRVNAITTVFVILAGLVWGIANPFMLNEQLTSTFIKYLPSYVRSVDA 136
 Query: 132 EANKLENEMLVSYKPOLQDMLTHTSQKALDYAQSPSKNAIDWAGNFAGAIARITVAIII 191
 + +KLL N+ L S++PQ+++ +T+ SOKA+DYA+ PSK A+ WAGNFA IAR+TVAIII
 Sbjct: 137 QVSKLLRNDLLASFRPQIENAVINFSQKAVDYAEPSPSKGAVTWAGNFASLIARVTVAIII 196
 Query: 192 SPFIIFYLRDSSSMKMGKLVNVLPLKLRVPMVRVLDINKQLSGYVQGGQVTVAVVVGPMF 251
 SPFI+FY LRDS MK V+ LP K+R P+ R+LGD+N+QL+GYVQ TVAI+VVGPMF
 Sbjct: 197 SPFIIFYLRDSSMKKEAFVSYLFTKMRQPIHRLIGDVRQLAGYVQSSSTVAIIVVGPMF 256
 Query: 252 SIMFSLVGLKYAITFGIAGFLNMIPYLGSLFLMIPVVMAMVQSPFMLVKVLWIFMIBQ 311
 SIMF++GL+YA+TFGIAGFLNMIPYLGSLFLA IPV I+A+V+GP +VKV ++F+EQ
 Sbjct: 257 SIMPTIIGLRVAVTFGIAGFLNMIPYLGSLFLATIPVFILALVSGSPVKVKVAVLPIVQ 316
 Query: 312 TIEGRFVAPLVGLNKLSTHPTITIMFLLTAGSMFGVGVFLPIPIYASIKVVKIKELFDNY 371
 TIEGRFV+PLVGL+KLSTHPTITIMF+LLTAGSMFGVGVFL IP+YAS+KVV+KE+P+MY
 Sbjct: 317 TIEGRFVSPVLVGLSKLSHPTITIMFILLTAGSMFGVGVFLGIPVYASIKVVKIKELFDNY 376
 Query: 372 KKVSGLYDEEVLVIEBKDHVK 393
 K +SGLY++E E+K VK`
 Sbjct: 377 KPISGLYKEE---EDIKKDVK 395

An alignment of the GAS and GBS proteins is shown below.

Identities = 243/389 (62%), Positives = 306/389 (78%), Gaps = 2/389 (0%)
 Query: 6 EKEFKNSLFFKILNMQAVIALMITFLNVLITPITKISPMFKVFDGLAVLLPLVLSG 65
 EK +SLF+KQ LNRQA +AL+IT L FLTIP+FTKISF+P PV F AV++LPLNIS
 Sbjct: 6 EKERTDLSFYKQFLNRQATVAILITLAFLLTIPVFTKISFLPMFVSPFVAVIMPLVIST 65
 Query: 66 ILYYLTKPMVTFLEKRGKIKRVTAIVLQVPTIILLLTMMSPTIMMSQRLHFMEDLPY 125
 +LYYL KE+V + G R T+I VF +I LL +HA+S F+PM+ QL F+EDLP Y
 Sbjct: 66 ILYYLTKPLVDLNLHGLPKRTTSITFVGLLTLFVWALSGLFVPMVQQLTSTFIEDLPKY 125
 Query: 126 VNKVQMETSSSFIDHNPWLKSYKGEISMLSHLSQVSYAEKFSKNIILWAGNLASTV 185
 V KV E + ++ N ML SYK ++ ML++ S +A+ YA+ FSKN +DWGN A +AR
 Sbjct: 126 VQKVKKEANKLLE-HEMLVSYKPOLQDMLTHTSQKALDYAQSPSKNAIDWAGNFAGAIAR 184
 Query: 186 VTVATIMAPFIFYLRDSSRMKNGFLMVLFTKLKQPTDIRLEMSNSQMSGVQGGIIVA 245
 +TVA I++PFILFY LRDS +MKNG + VLP KLR P R+L ++N Q+SGVQGG+ VA
 Sbjct: 185 ITVAIIISPFIFYLRDSSRMKNGLVNVLPLKLRVPMVRVLDINKQLSGYVQGGQTV 244
 Query: 246 ITVGIVPISIMYSIIGRLGVTLGLIAGVIANMVYLGSPVQIIPVFIALVAGPVMVVKVA 305
 I VG +FSIM++GL+Y +T GIAG LNM+PYLGSF+A IPV I+A+V GP M+VKV
 Sbjct: 245 ITVGPMFSIMFSLVGLKYAITFGIAGFLNMIPYLGSLFLMIPVVMAMVQSPFMLVKVL 304
 Query: 306 IPVFIQTILGRFVAPLVGLNKLSTHPTITIMFLLTAGSMFGVGVFLPIPIYASIKVVK 365
 ++F+TQT+IGRFV+PLVGLNKLSTHPTITIMF+LLTAGSMFGVGVFL IPYAS+KVV+
 Sbjct: 305 VIFMIBQITIEGRFVAPLVGLNKLSTHPTITIMFLLTAGSMFGVGVFLPIPIYASIKVVK 364
 Query: 366 KELFDMYKAVSLTVDV-VTEERSERVK 393
 KELFDMYK VSGLY +V V EE + VK
 Sbjct: 365 KELFDMYKVSGLYDEEVLVIEBKDHVK 393

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 888

A DNA sequence (GBSx0942) was identified in *S. agalactiae* <SEQ ID 2697> which encodes the amino acid sequence <SEQ ID 2698>. Analysis of this protein sequence reveals the following:

Possible site: 58
 >>> Seems to have no N-terminal signal sequence

-978-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2715(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9891> which encodes amino acid sequence <SEQ ID 9892> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA25160 GB:L16975 ORF1 [Lactococcus lactis]
 Identities = 132/345 (38%), Positives = 203/345 (58%), Gaps = 3/345 (0%)

Query: 79 INLAQIVAEVDGDIQAFVLYEYISDSQSYVSALLVMADLYDMBGLTVAREKILLASK 138
 +NLA+I +G+++A YL I + + Y+NL+ +ADLY E + A KL A +L
 Sbjct: 1 VMLAEIADNGNLDRALNYLQIFVNDENTIAALKIADLYQFEVDFETAIKLEAREL 60

Query: 139 SDDPLVTFGLAEHMLSLHNYQEAIRGYASLNRRIETTGVTYQRIKGSYAIMKFDPA 198
 SD PL+TF LAE Y AI YA L R+IL T +S YORIG SYA +G F+ A
 Sbjct: 61 SDDPLITFALARSYFQGNYSAAITVYAKLSERKILHETKISYQIGDSYQIGPENFA 120

Query: 199 IEFLEKAVDIEYDDLTVPFELATILYDQREYQKANLYFKQLDTINDFAGVEYIYXGSLRE 258
 I FLEK+++ T++++A + + +A FK+L+ ++ +F YE Y +L
 Sbjct: 121 ISFLEKSLFDEKPEPTLYKIALYGETHNETRALNFKLEKADVDFNLVYELAYATLEA 180

Query: 259 EHKSEALRLVQGGIRKNSFDQQLLLASQSLYELHNVHSSSEYLYKQARKVSENQDEIVM 318
 + + AL + ++G+KN LL AS++ ++L D ++E YL A + E DE V
 Sbjct: 181 NQSFKAALAKKGMKKNPNAVPLHFAKICFPLEMDKAAAEYLYVADNLPELHETVTF 240

Query: 319 RLNLVLEERFEFVLELDN-DNLNENILAKMNIKAHKALEMDSDVD--YYQSLYNDLKD 375
 L-NLY EE FE V+ L+ E++LAKW A AHKAE D Y + + +L +
 Sbjct: 241 LLNLVFNEDFEAVINLEELLEDEHLAKMLFAGAHKALENDEANALYELQLINLSE 300

Query: 376 NPEFLQDYAYILREFGYLDKQEVGKYLKLVDDIMESEWNNI 420
 NPEFL+DY L+E G + K + + +YL+VDD M + ++
 Sbjct: 301 NPEFLRDYIDFLKEIQISKTEPIIEQVLELVDDENMNL+TDL 345

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2699> which encodes the amino acid sequence <SEQ ID 2700>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2991(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 267/409 (65%), Positives = 336/409 (81%), Gaps = 1/409 (0%)

Query: 13 MNSKMIQVSIQMDLGHANKYFKALKNDPKEVLLLELGAYLESIGFLPQAKRYLDQIRP 72
 MNSKMI S+ QDL HA KYF+KALK D + L+ LG YLESIGFLP AKR+Y Q+
 Sbjct: 7 MNSKMIASLDQDLAHAKYKPKALKRDDASLIALAYLESIGFLPHAKRYLYLQAD 66

Query: 73 NYPEVAINLAQIVAEVDGDIQAFVLYEYISDSQSYVSALLVMADLYDMBGLTVAREKL 132
 +YF+ INLAQI AED IR+AFVLYD +S+DS Y-SALLVMADLYDMBGLT+VAREKL
 Sbjct: 67 DYPRNLINLAQIAEDDAIEAFVLYDQVSKDSPNYLSALLVMADLYDMBGLT+VAREKL 126

Query: 133 LKASKLSDDPLVTFGLAEHMLSLHNYQEAIRGYASLNRRIETTGVTYQRIKGSYAIM 192
 L A +S +PLV FGLAE++SL+H++BAI+ YA LDRN+ILE TG+STYORIG++YA +
 Sbjct: 127 LQAGVISPEPLVTFGLAEHMLSLHNYQEAIRGYASLNRRIETTGVTYQRIKGSYAIM 186

Query: 193 GKFDAAIEFLEKAVDIEYDDLTVPFELATILYDQREYQKANLYFKQLDTINDFAGVEYIY 252
 GKFDAAIEFLEKAV IEY+D TVFELAT++YDQE YQKANLYFKQL+TINDF+ GREY Y

-979-

Sbjct: 187 GKPEAAIEFLEKAVAIKYSDEIVFRLATIMYDQERNYQKANIYFKQLEITINPDYIPGYEGY 246

Query: 253 GLSLREHHSERARLRVQQGIRKNSFDQQLLLASQISYELMDVHSSESYLAKQAEKVSSEN 312
 LSL REHK+ RALRIVQQG+RKN+FD QLLLASQISYELMD ++E+YL QA++V+ +

5 Sbjct: 247 ALGLHEEHKTSERLRVQQGLRKNAFDSQQLLLASQISYELMDRQANRYLQAKAEVAVD 306

Query: 313 QDEIVWRISNLVLEERPFKEVLNDNLENTLAKWNIKAHKAHMD--SVDIYQSLYN 371
 +KI+MRL LY + ERPERV+ L+ + ++N+L KW IAKA+ ALE ++ ++ Y +

10 Sbjct: 307 DERILMLVLTLYFDARPEREIVALENRETFIDNVLTKWTLAKAYHALQEEVALALYNEISA 366

Query: 372 DLKONPEFLQDYAYILRFGFLDKQKVGKAYLKIVPDDIRMSRWNNI 420
 DL +NPEFLQDYAY+LREFG KA ++ AYLA+ VDD+ M ++++I

Sbjct: 367 DLAKNPEFLQDYAYILRFGQFKAIQMATYLRQVDDVNMQDFLDHI 415

- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 889

- A DNA sequence (GBSx0943) was identified in *S. agalactiae* <SEQ ID 2701> which encodes the amino acid sequence <SEQ ID 2702>. This protein is predicted to be alpha-acetolactate synthase (ilvK). Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2105 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 30 >GP:CAA01700 GB:A23961 alpha-acetolactate synthase [Lactococcus
 lactis]
 Identities = 396/559 (70%), Positives = 466/559 (82%), Gaps = 8/559 (1%)
- 35 Query: 4 SHNQYGADLIVDSLNDHVKYVFGIPGAKIDRVFDZLE-DKGPELIVARHEQNATFMAQA 62
 S Q+GA+L+VDSLNDH VKYVFGIPGAKIDRVFD LE ++GP++V RHEQ A FMAQA
 Sbjct: 2 SEKQFQANLIVDSLNDHVKYVFGIPGAKIDRVFDILENEEGPQMVVTRHEQAAPFMAQA 61
- Query: 63 VGRITGEPGVVATSGPGISNLATGLVATDEGDVALEIGGQVKRGDLKRAHQSMNVVA 122
 VGR+TSGPGVV+ TSGPG+SNLAT L+TAT EGD+LAIGGQVKR D LKRAHQSM+N
- 40 Sbjct: 62 VGRITGEPGVVVTSQPGVSNLATPLLTATSEGDAILAIGGQVKSRDLKRAHQSDNNG 121
- Query: 123 MLEPITIKYSAEVDHPNTLSEITANAYRLAKSGKFGASFISIPQDVTSFVSVAIKPLSA 182
 M++ TIKYSAEV DPNTLSE++ANAYR+AKSG PGA+F+SIPQDVTD+ VS+KAI+PLS
- 45 Sbjct: 122 MMQATKYSAEVLDPNTLSEITANAYRLAKSGHPGATFISIPQDVTDVTSKAIQPLSD 181
- Query: 183 PKLGSAVLIDINLYAQAINNVAFLVLLNGNASSGVTAAVRRLDAVKLPVVTFCQAG 242
 PK+G+AS+ DINLYAQA INVAFLVLLNGNASSGVTAAVRRLDAVKLPVVTFCQAG
- Sbjct: 182 PKMGASIDIDINLYAQAINKVAFLVLLNGAGSADKVAASSRLILTHVNI PVVETFCQAG 241
- 50 Query: 243 IVSRELESDTFFGRVGLFRNPQGMILKRADLVIAIGYDPIEYARNWNAKISARIIVID 302
 ++S +LE TF+GR+GLFRNPQGMILK+DLVIA+GYDPIEYARNWNAEI +RIIVID
- Sbjct: 242 VISHDLS-HTFYGRIGLFRNPQGMILKRSDLVIAVGYPDPIEYARNWNAEIDRIIVID 300
- 55 Query: 303 VEQAEIDTTFQPERELIGDMANTLIDLLPAIKGYELPESGKEYLKGIRNNIENVSVKPD 362
 AEIDTTFQPERELIGD+A TLD LEP++GY++P+G+K+YL GL E +FD
- Sbjct: 301 NATAEIDTTFQPERELIGDIAATLNLFLPAVRGYKIPKGTQDYLDGLH---EVAEHEHD 357
- Query: 363 RDSA-HGVRLPLDLDVLQENTTDMVTVDVVGSHYIMWARYFKSYEARHLLFNGNMQTL 421
 ++ G +HPLDL+ QE DD TTVVDVGS YIMW+FKSYE RHLFNGNMQTL
- 60 Sbjct: 358 TENTEGRHGLDLVSTFCBIVKDDSTVTVDVGSLYIMWAEHFKSYEPHLLFNGNMQTL 417

-980-

Query: 422 GVALPWAIISAALLRPNTKVISVSGDGGFLPSAQELSTAVRLHLPTVHIIWNDDGYMVEF 481
 GVALPWAI+ALLRP KV S SGDGGFLP+ QELSTAVRL+LPTIV IINWDD Y+M+P
 Sbjct: 418 GVALPWAITAALLRPGKKVSHSGDGGFLFTQSELSTAVRLNLPTVQIINWDDGYMVKP 477

5 Query: 482 QEEMKCYGRSSGVDPGPVDFVKYAESPGAKGYRVDSKDSFETLKQALIDAENGPIVLIDVP 541
 QEEMKCYGRS+ VDPG VD+VKYAE+ AKGYR SK+ E LK I GPV+IDVP
 Sbjct: 478 QEEMKCYGRSAVDPGYVDYVKYAEAMRAKGYRAHSKEELAEILKS--IPDOTTGPVVIDVP 535

10 Query: 542 IDYKDNVTGLGTFILEDEFY 560
 +DY DN+ L E +LP+EFY
 Sbjct: 536 LOYSDNIKLAELKLPDEFY 554

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 15 vaccines or diagnostics.

Example 890

A DNA sequence (GBS0944) was identified in *S.agalactiae* <SEQ ID 2703> which encodes the amino acid sequence <SEQ ID 2704>. This protein is predicted to be alpha-acetolactate decarboxylase (aldC). Analysis of this protein sequence reveals the following:

20 Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3095 (Affirmative) < succ>
 25 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9889> which encodes amino acid sequence <SEQ ID 9890> was also identified.

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAAS7941 GB:X82620 alpha-acetolactate decarboxylase [Lactococcus
 lactis]
 Identities = 139/239 (58%), Positives = 187/239 (78%), Gaps = 3/239 (1%)

35 Query: 16 MSETVKLFQYSTLSLHAGLYKGSITIGELLTHGDLIGITVHMIDGELIVLDGKAYQIG 75
 MSE +LFOY+TL +IMAGLY+G+TIGELL HGDIGIT+ IDGELIVLDGKAYQ
 Sbjct: 1 MSETITQLFQYNTLGAMAGLYEGTMTIGELLHGDIGITGLDSIDGELIVLDGKAYQ-- 58

40 Query: 76 TDGKAEITQLSDDVTVPYAAVLPHHIQKQFDINARINDKOLESMILKDFSGQNLFSKIKI 135
 G I+L+LD+ VPYAAV+PH + F + +K+LE+ I F+QGNLF+S+KI
 Sbjct: 59 -KGDKTIVELTDDIKVPYAAVPHIQLEVVFKQKFTVSDKLEDRIRSYFDGQNLFSIKI 117

Query: 136 KGTFSRMHVRMIPKSPQHKRFADIASNQPEPTRIRNVSGTLVGWITPELPHGVGKGFHVH 195
 G F +MHVRMIP++ +F +++ NQPE+T EN+ GT+VGWITPE+PHGV V G+H+H
 45 Sbjct: 118 TGKFFKMHVRMIPRAKSGTKFVEVSNQPEYTEENIKGTIVGWIWTPMPEPHGVSVAGYHLH 177

Query: 196 FISDDELTFGGHVMYDLSLTQCKVEIGKVDQLQCFFPTQDQCFKANKFDLQKREDIDLS 254
 FIS+D TFGGHV+D+ + G VEIG +DQL+Q PP QD++FL A+ D++ L++DID++E
 50 Sbjct: 178 FISDEDFTFGGHVLDFTIINGTVVEIGALDQLNQSFPVQDRKFLFADLDIALKKDIDVAE 236

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 891

A DNA sequence (GBSx0945) was identified in *S. agalactiae* <SEQ ID 2705> which encodes the amino acid sequence <SEQ ID 2706>. This protein is predicted to be fibronectin-binding protein-like protein A. Analysis of this protein sequence reveals the following:

```

5       Possible site: 57
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.5042 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15      >GP:CAA46282 GB:X65164 fibronectin-binding protein-like protein A
      [Streptococcus gordonii]
      Identities = 392/550 (71%), Positives = 462/550 (83%)

      Query: 1   MSPDGFLLHLLTNEQLQEIGRIQKVNQFPDHELVLITIRNNRNYKLLLSAHPVVGRIQ 60
      MSPDGFLLHR+T EL+ ++ GRIQK+NQFF+ ELVL IR+NR++ KLLLSAH VVG+VQ 60
20      Sbjct: 1   MSPDGFLLHMTTELRHLELVGGRIQKINQFPQELVLQIRSNRKLKLLLSAHSVVGVCV 60

      Query: 61   TTEANFQNPQNPNTPTMIMRKYLQGAIVETIQQIENDRILSIVSNKNEIGDHKATLVV 120
      T+ F+NP PNTF M+MRKYLQGAIVE IQQ+ENDRILEI VSNKNEIGD + TLV+ 120
25      Sbjct: 61   LTDITFENPAVPNTFTVMVRKYLQGAIVETIQQIENDRILEISVSNKNEIGDSVAVTLVLT 120

      Query: 121  EIMGKHSNIIILDKNSHKIESIKHVGFSQNSYRTILPGSTYIAPPKTKAINPFDISDQT 180
      EIMGKHSNII+DK KIIE+IKHVGFSQNSYRTILPGSTY+APP+T ++NPF + D+ 180
30      Sbjct: 121  EIMGKHSNIIILDKASGKIIRAKHVGFSQNSYRTILPGSTYVAPPQTGSLNPFVIGVEK 180

      Query: 181  LFEILLQINDLSPKNLQQLLQGLGRDITALESLHCLKDNKLNDFRQFSREYYPSLTKSFS 240
      LFE+LQT ++ PK L Q+ OGLGRDIT ELS L ++L FR FF+ PSITKESFS 240
35      Sbjct: 181  LFEILLQETIEPKRLQIFQLGLGRDITALESGRLTTRDLRTKRAFFASPTQPSLTKSFS 240

      Query: 241  AVQSSSHETPQSGLLDYYQKAEKDRIAQAASDLHRVQSELEKNIKKLAKQDEEL 300
      A+ FS S +L +LLD +Y++KAE+ R+ QQAS+LI RV++ELSKN KKL KQ+ DEL 300
40      Sbjct: 241  ALVFSDSKTMSTLSLELLDTFFYDKAERYVRVQASSELIRRVENELEKNRKLKGQDEEL 300

      Query: 301  LATENAEFRCKGELLTYLMSVPMNQVVLNDYNTTQTEISLDRALTPQNAQRYFK 360
      LAIE AEEFRCKGELLT+L VR+QD V LDNYT + I L-ID+ALTPNQNAQRYFK 360
45      Sbjct: 301  LATENAEFRCKGELLTYFLHQVFNDDQVELNDYNTTGKILITLDKALTPQNAQRYFK 360

      Query: 361  RYQKLKEAVKHLKGIISDTENVTITLYSSVETSINHASMDINDIRELVEVTFIKRRAHD 420
      +YQKLKEAVKHL +I +T TI YLSSVET+L AS+ +I +IKREL++TQFT+RR + 420
50      Sbjct: 361  RYQKLKEAVKHLTSLIRESTRITITLYLSEVETALQASLTSIAEIRBELIQTFIRNRQRE 420

      Query: 421  KQHKKRKPQYLASDGKTIINVGRRNLQNDLTFPMARKGELMFAKADIPGSHVLIRNLI 480
      K KRRKPE+YLASDG+TII+VGRNMLQNDLTFPM+K ELAFKADIPGSHV+I NL 480
55      Sbjct: 421  KIQRRKKPEYLASDGQITILVGRNMLQNDLTFPMAKKOSLWFAKADIPGSHVVTIGNL 480

      Query: 481  NPSDEVKTDAAELAAAYSKARLSNLVQVDMIEAKKLKPGSTKPGFVYTTGQKTLRVPTP 540
      PSEDEVKTDAAELAAAY+SKARLSNLVQVDMIE KKLKPK+G KRGFVYTTGQKTLRVPTP 540
60      Sbjct: 481  QPSDEVKTDAAELAAAYPSKARLSNLVQVDMIEIKKLKPKTGKPGFVYTTGQKTLRVPTP 540

      Query: 541  QSKIDSLKIK 550
      +KI S+K++
65      Sbjct: 541  ADKIKSMKIQ 550

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2707> which encodes the amino acid sequence <SEQ ID 2708>. Analysis of this protein sequence reveals the following:

```

60      Possible site: 38

      >>> Seems to have no N-terminal signal sequence

```

-982-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.5434 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein differs significantly from L28919 in its mid-region:

10 Query: 223 QHPQGLGRDTAKELALLTDT
 F L + T K + ELLTDT
 Sbjct: 121 PAPSRLLRGSTPKRIGELLTDT

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 421/549 (76%), Positives = 487/549 (88%)

Query: 1 MSFDGFFLHHLTNELQBIQKRIQKVNQPPDHELVLTIRNRRNRYKLLLSAHPVGRIG 60
 MSFDGFFLHHLTNEL+E + GRIQKVNQPP+ ELVLTIRN+R+NYKLLLSAHPVGR+Q
 Sbjct: 27 MSFDGFFLHHLTNELKENLGYRIQKVNQPPFERELVLTIRNRRNRYKLLLSAHPVGRVQ 86

20 Query: 61 TTEANQNPQNPNFTFMIRKYLQGAVIDTIQIENDRIELTVSNQNEIGDHIGKATIAV 120
 T+A+QNFQ PNTFTMIRKYLQGAVID ++QI+NDRI+EI VSNQNEIGD I+ATL++
 Sbjct: 87 ITQADQNPQVNPFTFMIRKYLQGAVIDQLQIENDRIIEIKVSNQNEIGDAIQATLII 146

25 Query: 121 EIMGKHSNIIILIDKNEHKIIESIKHVQFSQNSYRTILPGSTYIAPPKTKAINPFDISDQT 180
 EIMGKHSNII+L+D+ E+KIIESIKHVQFSQNSYRTILPGSTYI PPKT A+NPF I+D
 Sbjct: 147 EIMGKHSNII+LVDRAENKIIESIKHVQFSQNSYRTILPGSTYIIEPPKTAAVNPFITIDVP 206

30 Query: 181 LFEILQITNDLSKPNLQQLQLGRDTALELSHCLKDKNLNDFRQFSSREYTPSLITKSPS 240
 LFE+LOT +L+ K+LQO QGLGRDTA EL+ L +KL FR+FF+R +LT SF+
 Sbjct: 207 LFEILQITELTVKSLQQRHQLGRDTAKELALLTDTKLKRFRRFFARPQTQANLITASFA 266

35 Query: 241 AVQFSSSHETFOSLQGLDYYYQKAEKORIAQASDLIRHVQSELENNIKKLAQODEL 300
 V FS SH TF++L +LD++YC+KAE+DKI QQASDLIRHVQ+EL+KN KL+KQ+ EL
 Sbjct: 267 PVLPSDSHATFETLSMLDHPYQDKAERDINQQASDLIRHVQTELDNRNKLKQGAEL 326

40 Query: 301 LATENAEFRQKGEILLTYLSMVPNNQDVVVLQNYTNTQITIEISLDRALTQNAQRYFK 360
 LATENAE FRQKGEILLTYLS+VFNNDQ V+LQNYT + IEI+LD+ALTQNAQRYFK
 Sbjct: 327 LATENAEFRQKGEILLTYLSLVFNNDQSVILQNYTTEKIEIALDKALTQNAQRYFK 386

45 Query: 361 KYQKLAKEAVKHLKGIISDTENTITYLESVETSLNWSMEDINDIRELVBETGFIKRAHD 420
 KYQKLAKEAVKHL G-I+DT+ +ITY ESV+ +L+ AS++DI DIREEL +GF+K R D
 Sbjct: 387 KYQKLAKEAVKHLGLIADTKQSTITYPESVDYNLSQASIDDIIDIRELVBETGFIKRAHD 446

50 Query: 421 KQKRRKKPEQYLASDGKTIIMVGRNNLQNDLFTKMARKEGLWPHAKDIPGSHVILRDNL 480
 K+KRRKKPEQYLASDG TI+MVGRNNLQ+N+ELFTKMA+KGLWPHAKDIPGSHV+I+DNL
 Sbjct: 447 KRRKKPEQYLASDGTIIMVGRNNLQNEELTFMARKEGLWPHAKDIPGSHVITKDNL 506

Query: 481 NPSDEVKTDAAELAAAYSKARLNLVQVDMIRAKKLNKPSGTYKGFVTTYGKTLRVPT 540
 +PSDEVKTDAAELAAAYSKARLNLVQVDMIRAKKL+KPSG KGFVTTYGKTLRVPT
 Sbjct: 507 NPSDEVKTDAAELAAAYSKARLNLVQVDMIRAKKLNKPSGAKGFVTTYGKTLRVPT 566

55 Query: 541 QEKIDSLKL 549
 Q KI S+KL
 Sbjct: 567 QAKILSMKL 575

SEQ ID 2706 (GBS81) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 6 (lane 2; MW 64kDa) and in Figure 6 (lane 5; MW 64kDa). The GBS81-His fusion product was purified (Figure 190, lane 3) and used to immunise mice. The resulting antiserum was used for FACS (Figure 319), which confirmed that the protein is immunooaccessible on GBS bacteria.

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 892

A DNA sequence (GBSx0946) was identified in *S.agalactiae* <SEQ ID 2709> which encodes the amino acid sequence <SEQ ID 2710>. Analysis of this protein sequence reveals the following:

```

Possible site: 53
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL      Likelihood = -9.08      Transmembrane      6 - 22 ( 1 - 24)

----- Final Results -----
      bacterial membrane --- Certainty=0.4630 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAF94260 GB:A5004191 conserved hypothetical protein [Vibrio cholerae]
Identities = 111/295 (37%), Positives = 164/295 (61%), Gaps = 1/295 (0%)

Query: 36 QVVKIGILQYVTHDALDAIEKGVEDGLAQEGYK-GKKVLT/VLNAEADQSKIQAMSKQLV 94
+ K+ + Q V H ALDA +G+ DGL +GY+ GK ++ A+ + + +++Q V
Sbjct: 26 KTAQAVAVSQIVEHPALDATRCGLLDGLKAGGYEGKNLEFDYKTAQGNPALVAQIARFV 85

Query: 95 NHHNDILIGIATPSAQGLAASTKDTPIIMGAVSDPLGAKLVNMMKKPTTNTVGLSNVVPT 154
+ D+L+GIATP+AQ L ++TK PI+ AV+DP+GAKLV +++P NVTGLS++ P
Sbjct: 86 GENPDVLVGLIATPTAQALVSNATKTIPIVFTAVTDPVGAKLVKQLEQPGKNVTGLSDLSFV 145

Query: 155 KQTVQLIKDITPNIKRIGILYASSEDNSVSQVTEPTKYAQKAGLEVLYKVSVPSTNEIKTS 214
+ Q V+LII+I FN+K IG+Y+ E N+VS + A K G+++++ + +++++
Sbjct: 146 EQEVELIKEILLPNVKSIGVVYTPGRANVSMLELLKLSAAKHGIKLVEATALESADVCBA 205

Query: 215 MSVWTKKVDVAVFPQDNTIASAFKTVIVAANQANIFVYSSVDVMEQGSIASVASQSYGL 274
+ +K D ++ DNT+ASA +IVAANQA FV+ + + VE+G+IAS+ Y +
Sbjct: 206 TQALIAEKSDVIALINDNTVASAIEGMIVAANQAKTFVFGAATSVERGATIASLGFQFYQI 265

Query: 275 GLSTAKQAIKVLGKGVKVDVFKVVDITGKPSLNLIKAHGLGIKPKKINKQARIT 329
G++TA ++L GK + V+V ++N HA+ LGI IP+ ++ A T
Sbjct: 266 GVQTDYVAALIEGKPGSLDVQVAGKSDLVINKTAEGQLGITPEAVLARATST 320

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2711> which encodes the amino acid sequence <SEQ ID 2712>. Analysis of this protein sequence reveals the following:

```

Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL      Likelihood = -11.25      Transmembrane      6 - 22 ( 1 - 27)

----- Final Results -----
      bacterial membrane --- Certainty=0.5501 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:AAF94260 GB:AB004191 conserved hypothetical protein [Vibrio cholerae]
Identities = 103/304 (33%), Positives = 178/304 (57%), Gaps = 1/304 (0%)

Query: 17 VIGSLLSKGVSKENROLANQNTTIGILQFVTHEALDDIKRGIEDQLK-KQMPQKQNVVI 75
VI + + G + + + + Q V H ALD ++G+ D LK K + +N+
Sbjct: 6 VIATAVLAGAALLSSQSIMAKTAQVAVSQIVEHPALDATRCGLLDGLKAGGYEGKNLEF 65

Query: 76 KVMNARETSQSKIQTMRQLVQSGSDIVIGIATPAAQGLAATSKDIPVVMASVDFVGSRL 135
A+G+ + ++RQ V D++GIATP AQ L + +K IP+V +AV+DPVQ+L+
Sbjct: 66 DYKTAQGNPALVAQIARFVGENPDVLVGLIATPTAQALVSNATKTIPIVFTAVTDPVGAKL 125

Query: 136 VMQLDQPEANVTGLSNKVFVKQTIDMKKLTPHVKTFVGLIYASNRDNLSDQVKEFRRLAR 195

```

-984-

V QL+QP NVTKLS+ PV+Q ++L+K++ P+VK++G++Y E N++S ++ + A
 Sbjct: 126 VQLQEQGKGVTKLSDLSPVBCHEVLIKSLPANKSIGVVPNGEANAVALKLLKLSAA 185

Query: 196 KKGQVQISYAVSTNEVPATMSVMLGKVDVAFPIQDNTLASAPSSVMTSKAAKIPVYTS 255
 K G +++ + +V + + + K D ++ DNT+ASA ++ + AK PV+ +
 Sbjct: 186 KKGKLVTEALSKADVQSATQATAEKSDVIYALIDNTVASAIEGMIWAWQAKTPVFGA 245

Query: 256 VDRMVEKGLAASIQNQYDLGVQNTANQVYLKIKGKRVDPVPKQVVDIGQPLINQVAAEL 315
 VE+G +A++ + Y +GVOTA+ V +++GK + V+V +INK A +L
 Sbjct: 246 ATSYVFERGATASLGPDYVYQIGVQTDYVAAILBGEKPGSLDVQVAKGSDVINNTAAEQ 305

Query: 316 GIAL 319
 GI I
 Sbjct: 306 GITI 309

An alignment of the GAS and GBS proteins is shown below.

Identities = 181/322 (56%), Positives = 252/322 (78%), Gaps = 1/322 (0%)

Query: 1 MNKKGLIATLILLITLVVGLFYNK-SKKRNLSEKQVVKIGILQVYTHDALDAIKGVE 59
 MNKK LIATL++L++V+G L S++ +L+ +Q + IGILQ+VTH+ALD I++G+E
 Sbjct: 1 MNKKSLIATLILLITLVIVIGSLSKGVSKENRDLANQQNITIGILQVTHREALDDIKRGIE 60

Query: 60 DGLAQBSYKGGKVKLVINAEADQSKIQAAMSKQLVNHNDILIGIATPSAQGLAASKDT 119
 D L ++ + + V + V+NAE DQSKIQ MS+QLV +DI+IGIATP+AQGLAA++KD
 Sbjct: 61 DQKKGMPQKQNVVVKVNAEADQSKIQTMSRQLVQSGSDIVIGIATPAAQGLAATSDI 120

Query: 120 PIIMGAVSDPLGAKLVDMKKPTTNVGLSNVVPKOTVOLIKDITFNKIRIGILYASSE 179
 P++M AVSDP+G++LV + +P NVTGLSN VP KQT+ L+K +TP++K +GILYAS+E
 Sbjct: 121 PVVMSAVSDPVGSRVLVQDQPEANVTGLSNKVPVEQTIDMKKLTFFHVKTGVILYASNE 180

Query: 180 DNSVSQVTEFTKYAQAGLEVLYKVSVPSTNEIKTSMVMTXKVDVFPVQDNTIASAFRT 239
 DNS+SQV EF + A+K G +V+ Y+VPSTNE+ +MSVM KVDVFP+QDNTIASAF +
 Sbjct: 181 DNSLSQVKEPRLLARKKSGYQVSYAVPSTNEVPATMSVMLGKVDVAFPIQDNTIASAFSS 240

Query: 240 VIVANQANIPVYSSVDTMVEQGSIASVAGSQCYGLGLETAKQAIVLKGPKVDVVPKVI 299
 V+ + A IPYI-SVD MVE+G +A+++Q+CY LG++TA Q +K+++GK V DVPKVA+
 Sbjct: 241 VMTTSKAAKIPVYTSVDMSVMEKGLAALISQNYDLGVQNTANQVYLKIKGKRVDPVPKVV 300

Query: 300 DTGKPSLNKAAKHLGIKIKPK 321
 D G+P +N A LGI I K+
 Sbjct: 301 DIQQPLINKVAELGIAIKKE 322

SEQ ID 2710 (GBS254) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 43 (lane 4; MW 27kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 48 (lane 3; MW 59.6kDa).

GBS254-GST was purified as shown in Figure 203, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 893

A DNA sequence (GBSx0947) was identified in *S. agalactiae* <SEQ ID 2713> which encodes the amino acid sequence <SEQ ID 2714>. This protein is predicted to be probable permease of ABC transporter (rbsC). Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -15.12	Transmembrane	127 - 143 (119 - 151)
INTEGRAL	Likelihood = -8.81	Transmembrane	206 - 222 (200 - 222)
INTEGRAL	Likelihood = -6.46	Transmembrane	260 - 276 (258 - 287)
INTEGRAL	Likelihood = -5.84	Transmembrane	234 - 250 (231 - 257)

-985-

INTEGRAL Likelihood = -4.78 Transmembrane 55 - 71 (54 - 72)
 INTEGRAL Likelihood = -3.61 Transmembrane 177 - 193 (176 - 194)
 INTEGRAL Likelihood = -3.35 Transmembrane 84 - 100 (83 - 102)
 INTEGRAL Likelihood = -1.91 Transmembrane 10 - 26 (10 - 26)

----- Final Results -----

bacterial membrane --- Certainty=0.7050(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AG07224 GB:AB004801 probable permease of ABC transporter
 [Pseudomonas aeruginosa]

Identities = 116/288 (40%), Positives = 185/288 (63%), Gaps = 9/288 (3%)

Query: 2 IISVSQGLLWGLGLGIYITFRILKFPDMITBGSFPLGGAVCVTLMQGVNPFILATILG 61
 + ++ GL++ ++ LG+++FR+L+FPD+T +GSFFPLGGAVC TL+ G +P AT+
 Sbjct: 6 LFGALETIGLIFSLVALGVFISFRLLRFPDLTVGDSFPLGGAVCATLIALGWDPPYSATLAA 65

Query: 62 MLSGMLAGFVTGLLYTKKMPITLAGILVMTSCNSILMVMKRNGLNEITQLKDFLPF 121
 +G LAG TGLL K KI +LA IL+M + +SI L +M + N+ L TL L
 Sbjct: 66 TAAGALAGLATSLNLVKKIMDLLASILMMIALYSINLRKNGKPNVFLIARPTLFTLLQP 125

Query: 122 SNDNLNLVGLIALLWISA---LIYFLYTRLGQAYIATGDNFOMAKSFGIDTKMEMLG 178
 + + L+ + +VI+ A L +F T+ G A ATG NP MA++ G++T N +LG
 Sbjct: 126 EWLSDVFRPFLLLVFIVIAKLLLDWFFTTQKSLAIRATGNSFRMARAGQVNTGGNILLG 185

Query: 179 LIVSNGLIALSGALVSGQDGVADYVSKIGIVIVIGLASIIGB-VLYSTGUTLFRRIIAIV 237
 + +SN L+AL+GAL +Q G AD+S GIG IVIGLA++I+GE +L S L L +A++
 Sbjct: 186 MAISNALVALAGALFACTQSGADISMGITIVIGLAIVTGSILSRRLIL--ATLAVI 243

Query: 238 VGSILYQPLITAVI---ALGPNYVILKLSAIVLIGICLAVFLATKIL 282
 +G+I+YF I + +Q L L +A++ + L+P+K +L
 Sbjct: 244 LGAIYVRFFIALANSDFTIGLQAQDLNLVAVLVIVLVALVIMPMKKRL 291

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2715> which encodes the amino acid sequence <SEQ ID 2716>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -10.46 Transmembrane 131 - 147 (125 - 156)
 INTEGRAL Likelihood = -8.65 Transmembrane 210 - 226 (204 - 230)
 INTEGRAL Likelihood = -8.17 Transmembrane 265 - 281 (261 - 283)
 INTEGRAL Likelihood = -7.22 Transmembrane 238 - 254 (233 - 261)
 INTEGRAL Likelihood = -3.03 Transmembrane 89 - 105 (87 - 107)
 INTEGRAL Likelihood = -2.60 Transmembrane 63 - 79 (62 - 79)
 INTEGRAL Likelihood = -2.23 Transmembrane 180 - 196 (180 - 198)
 INTEGRAL Likelihood = -2.13 Transmembrane 14 - 30 (14 - 30)

----- Final Results -----

bacterial membrane --- Certainty=0.5182(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AG07224 GB:AB004801 probable permease of ABC transporter
 [Pseudomonas aeruginosa]

Identities = 118/285 (41%), Positives = 186/285 (64%), Gaps = 7/285 (2%)

Query: 6 IISVSQGLLWGLGLGIYITFRILNFPDMITBGSFPLGGAVAVTAISLGNPFSLTILG 65
 + ++ GLI+ ++ LG+++FR+L FPD+T +GSFFPLGGAV T I+LGP+P +TL
 Sbjct: 6 LFGALETIGLIFSLVALGVFISFRLLRFPDLTVGDSFPLGGAVCATLIALGWDPPYSATLAA 65

Query: 66 MLSGMLAGFVTGLLYTKKMPITLAGILVMTSCNSILMVMKRNGLNEITQLKDFLPF 125
 +GALAG TGLL K K+ LLA IL+M + SI L +M+ N+ L + L

-986-

Sbjct: 66 TACGALAGIATGLINVKLLKINDLLASTILMMIALYSINLRIMGKPNVFLIARPTLFTLLQP 125

Query: 126 SIDLNSLI/TGLITVIVIVIS---VLIVFLYTNLQAYIATGUNKMAKSPGINTDWMVEVVG 182
 + + L+ V IVI+ +L+P T G A ATG N MA++ G+NE M ++G

5 Sbjct: 126 EMLSYVFRPLLLVFTVIAAKILLDFWFTTQKGLAIRATGSGNFMARAJAGVNTGGMILLG 185

Query: 183 LVVSNISLIALSGALVSCQDGYADVSKGIGVIVIGLASTIVGEVLYSTGLTLERLIAIVI 242
 + +SH+L+AL+GAL +Q G AD+S GIG IVIGLA++IVGE + + +L L A++

10 Sbjct: 186 MAISNALVALGALFACQSGADISNGIGIVIGLAADVIGESILPERRLILATL-AVIL 244

Query: 243 GSILYQPLISVWIT--LGFNTSYKLISALVIALCLMIPVVKER 284
 G+I+Y+P I++ + +G L L+R++ + L+IP++K+R

Sbjct: 245 GAIVYRFFTAIALNSDFIGLQAQDLNLVAVLVTVVALVIMPMKKR 289

15 An alignment of the GAS and GBS proteins is shown below.

Identities = 227/287 (79%), Positives = 259/287 (90%)

Query: 1 MIISVSQQLMWIGLIGLYLTFRILKFPDMTEGSPFLGGAVCVTMMNQVNPILATIL 60
 MIISVSQQL+WG+LGLGILYLTFRIL FPCDMTEGSPFLGGAV VT ++ G NP L+T+L

20 Sbjct: 5 MIISVSQQLTWIGLIGLYLTFRILNFFDMTEGSPFLGGAVAVTAISLGMNFFLSTLL 64

Query: 61 GMLSGMLAGFVTGLLYTEGKIPTILAGILVMTSCHSIMLAVMKRANLGEIQTLDKFLP 120
 GMLSG LAGF+TGLLYTEGK+PT+LAGILVMTSC+SIMLAVM RANLGL++ + ++D LP

25 Sbjct: 65 GMLSGLAGFVTGLLYTEGKMPITLILAGILVMTSCSIMLAVMKRANLGLDHRKIQDCLP 124

Query: 121 FSDILNLLVIGLIAILLVISALIYFLYTRLGQAYIATGDNPMKSPGIDTDKIMENGLI 180
 FS DEN L+ GLI ++VIS LIYFLYT LGQAYIATGEN DMKSPGI+TD ME++GL+

30 Sbjct: 125 FSDILNLLTGLITVIVIVISVLIVFLYTNLQAYIATGUNKMAKSPGINTDWMVEVGLV 184

Query: 181 VBNGLIALSGALVSCQDGYADVSKGIGVIVIGLASIIIGEVLYSTGLTLERLIAIVGS 240
 VSN LIALSGALVSCQDGYADVSKGIGVIVIGLASII+GEVLYSTGLTL ERLIAIV+GS

35 Sbjct: 185 VSNLIALSGALVSCQDGYADVSKGIGVIVIGLASIIIGEVLYSTGLTLERLIAIVGS 244

Query: 241 ILVQFLITAVIALGFNTNYLKL+SAIVLIGICIMVFLKTKILKGVRL 287
 ILVQFLI+ VI LGFNT+YIKL SA+VL +CLM+PV+K + HGVRL

Sbjct: 245 ILVQFLISVITLGFNTSYKLISALVIALCLMIPVVKERFFHGVRL 291

A related GBS gene <SEQ ID 8681> and protein <SEQ ID 8682> were also identified. Analysis of this protein sequence reveals the following:

40 Lipop: Possible site: -1 Crend: 0
 Mod: Discrim Score: 4.24
 GVH: Signal Score (-7.5): -6.43
 Possible site: 24

45 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 8 value: -15.12 threshold: 0.0

Feature	Likelihood	Transmembrane	Count	Range
INTEGRAL	-15.12	127 - 143	(119 - 151)	
INTEGRAL	-7.54	206 - 222	(201 - 225)	
INTEGRAL	-6.48	260 - 276	(258 - 282)	
INTEGRAL	-5.84	234 - 250	(231 - 257)	
INTEGRAL	-4.78	55 - 71	(54 - 72)	
INTEGRAL	-3.61	177 - 193	(176 - 194)	
INTEGRAL	-3.35	84 - 100	(83 - 102)	
INTEGRAL	-1.91	10 - 26	(10 - 26)	
PERIPHERAL	4.77	36		

55 modified ALOM score: 3.52

*** Reasoning Step: 3

----- Final Results -----

60 bacterial membrane --- Certainty=0.7050(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF00338 (298 - 1146 of 1461)
 GP|9950013|gb|AA070224.1|AE004801.2|AE004801 (4 - 291 of 296) probable permease of ABC
 transporter {Pseudomonas aeruginosa}

```
%Match = 20.2
%Identity = 40.8 %Similarity = 68.3
Matches = 116 Mismatches = 84 Conservative Subs = 78
```

126 156 186 216 246 276 306 336
YGLGLETA RQAIVLHGKPKVDVPVKVIDTGKPSLNLKAAKHLGIKIPKKIMKQAIEITVKVDDAKKGFMIISSVSQGLLW
| : : : || :
MSLPSLFGALIEGLIF
10

[illegible][illegible][illegible]

```

1077      1086      1116      1146      1176      1206      1236      1266
VIALGFNTNY-----LKLPSAIVLGIICLAVPVLTKILRGVRL*W**KB*S*KKQPYKSVWV*QK*KRY*IMLI*VFM
| | : : : | | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
--ALALNSDFIGLQAQDINLVTAVLVTVLPMKKRLGKKGGA
      270      280      290

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 894

A DNA sequence (GBSx0948) was identified in *S. agalactiae* <SEQ ID 2177> which encodes the amino acid sequence <SEQ ID 2178>. This protein is predicted to be ABC transporter (potA). Analysis of this protein sequence reveals the following:

Possible site: 36
>>> Seems to have an uncleavable N-term signal seq

```

----- Final Results -----
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9887> which encodes amino acid sequence <SEQ ID 9888> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF86640 GB:AF162694 ABC transporter [Enterococcus gallinarum]
Identities = 117/252 (46%), Positives = 167/252 (65%)

Query: 19 MVMKIIELKEATVQVSNGLAEMKTIIDHVNLSIYEHDFITILGGNGAGKSTLPNVIAGTL 78
M + + + + G + L + + L + + DEITI+GGNGAGKSTI N IAGT+

-988-

Sbjct: 1 MTPFVLTISDLHQTFEKGTTINENHVLRSGLDITWNSCDFITTIQGNAGKSTLNSIAGTI 60

Query: 79 MLSSGNIYIMQDVTHLSAEKRAKYLRSVPQDPKMGTAPEMTVAENLLIAKFRGEKRPVLV 138
G I + ++T E +R+K +SRVPQDP+KCTA R+TV ENL +A RG+ R

5 Sbjct: 61 PTEQQKIVILGDKETTRHSVTRRSKEISRVFPDPRMGTAVALRVTSSENLALAYIKRQVRGFS 120

Query: 139 PRKIINYTEFQKLIARTNGNLDHLETPITGLLGGCQQAELSLLMATLKKPNLLILDEHT 198
+ F++ +AR QI+ L T GLISGCGQRA++LIMNTL++P L+LIDRHT

10 Sbjct: 121 SGVKGKIRAFFKEKLARINLGLNRLITTEIGLISGCGQQAELTILMATLQKPLILDEHT 180

Query: 199 AALDPRTSVISGLTDEFTIKQDSLTALMITHHMDALKYGNRVLVMDKGIYVRDLNQAQK 258
AALDP+TS+++M LTD+ I++ LTA M+TH MEDA++YGNR+++ GKIV D+ +K

15 Sbjct: 181 AALDPKTSMTVMALTDQLIQSQQLTAPFVTHMEDAIRYGNRLIMLQKSIIVDITGEK 240

Query: 259 NKMLADTYQLF 270
+ + D LF

15 Sbjct: 241 QSLTVFDLMALF 252

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2719> which encodes the amino acid sequence <SEQ ID 2720>. Analysis of this protein sequence reveals the following:

Possible site: 58
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2249 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 186/250 (74%), Positives = 210/250 (83%)

Query: 22 KIIELEKATVVOVSNGLAEKMTILDHVNLSTYEHDFITLGGNGAGKSTLHVIAQTMLG 81
KIIELE ATV V NG + KTIID+V L+IYEHDF+TILGGNGAGKSTLHVIAQT L+
35 Sbjct: 3 KIIELEINATVDVINGFEDAKTILDNVTITIEHDFITLGGNGAGKSTLHVIAQTMLG 62

Query: 82 SMYIYIMQDVTHLSAEKRAKYLRSVPQDPKMGTAPEMTVAENLLIAKFRGEKRPVLVPRK 141
G I I+QCDVT+ AEKRA YLSRVFPD KMGTAPEMTVAENLL+R+ RG KR L RK
Sbjct: 63 RQRIILQDVTHMFAEKALYLSRVFPQDSKMGTAPEMTVAENLLIARQGGKRSLSRK 122

40 Query: 142 IINYTEFQKLIARTNGNLDHLETPITGLLGGCQQAELSLLMATLKKPNLLILDEHTAL 201
+ + F+ L+ RTNGI++HLETP GLISGCGQQAELSLLMATLKKP LILLDEHTAL
Sbjct: 123 ITEHLASFEDLVKRTGNGLKHELETPAGLISGCGQQAELSLLMATLKKPALLILDEHTAL 182

45 Query: 202 DPTSTSVISGLTDEFTIKQDSLTALMITHHMDALKYGNRVLVMDKGIYVRDLNQAQK 261
DP+TS SLN LTDEF+ +D L+TALMITHHMDAL YGNR++VMKDGI I++DLNQ +K ++
Sbjct: 183 DPKTSQSLMQLTDEFTIKQDSLTALMITHHMDALTYGNRLIVMDKQNIITKIDKQMEKQL 242

Query: 262 AIADYQLFD 271
I DYQLFD

50 Sbjct: 243 TITDYQLFD 252

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 895

55 A DNA sequence (GBS0949) was identified in *S.agalactiae* <SEQ ID 2721> which encodes the amino acid sequence <SEQ ID 2722>. Analysis of this protein sequence reveals the following:

Possible site: 33
>>> Seems to have no N-terminal signal sequence

60 ----- Final Results -----

-989-

bacterial cytoplasm --- Certainty=0.1930 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 RGD motif: 415-417

The protein has homology with the following sequences in the GENPEPT database.

>GF:BA06117 GB:AP001515 unknown conserved protein [Bacillus halodurans]
Identities = 236/549 (42%), Positives = 362/549 (64%), Gaps = 2/549 (0%)

10 Query: 4 IKMALGSGVRENGNLYVVEVNDIFVLDAKLKYFENEQGVDDVVPINLQYLIENKKRVQ 63
I++ ALGKV R GKN+YVVEV+D +V+DNL +P+E LGDVVVP+ ++ YL+EN++RV+
Sbjct: 9 IRVFALGSGVEIGKQWYVVEVDDLFVIDGLMPDDEMLGVDDVVPIDISYLVNEDEVR 68

15 Query: 64 GIFLTHGHADAIGALPYLIAHVEAPVFGSELTIELAKLFVNSTAVKKFNNFHVDSITE 123
I LTHGH D IG LPY++ ++ PV+G+ LT+ L + +K + ++ +IDS +
Sbjct: 69 AILLTHGHEDHIGELPYVLQKLVFPVYGTGLTLGLVEEKLKEAGLIRSAK-LKLLDSHSR 127

20 Query: 124 IEPQDAVISFPFKTHSIPESMOIVIGTGRGNIVYTGDPKFDQAAKYQYTDLARLAIGR 183
++ +SPF+T HSIP+S+GI I T +G IV+TGDPKFDQ Q ++ ++A IG
Sbjct: 128 LKLGSTFVSFPFRTHSIPDSVGICQTSQGFIVHTGDPKFDQTFPDKQAEIGKQAAIGH 187

25 Query: 184 DGVLLALLSDSANATSNBQVASEYEVGDEIKSVIEDAAGRVIVAVASNLIRIQQVDDAA 243
GV L L L S D S N A S E E V G I E + G R + I V A S N + R + Q Q V A A
Sbjct: 188 KGVLLCLSDSTNAERPGMTKSETEVGRGIAEPEQTGRIVITPASNVHRQVQVHAJ 247

30 Query: 244 ENGRRVVLTFGPDINIVRTAIPMKRIHIADENMILPKDMTRYEDNELLIETGRMGPEI 303
R++ + G + +V A R+ + D+ + I ++++Y+D + I+ TG GEP+
Sbjct: 248 ATNRKLAVAGRSWVKVUSIAERLGLYLEAPDD-LFID:EEVSKYDDERVAIITTSQSEPM 306

35 Query: 304 NGLQKQVIGRHRHYVQIKDGLVFIPTTSLAKAVVARVENLYKAGSVKLITONLRAV 363
+ L +MA G HR + I + D V I TP E V+ + +L++ G V + + S
Sbjct: 307 SALSMAKGAHRQITITENTVILIAATPIGNERSVSTIVOLLHRIGADVIFGHSKVIAS 366

40 Query: 364 GHANGRELQMLNMLKPKYLFPIQGEYRDLASAHAGLAQVGSADDIYIVKRGDINVLEK 423
GH + EL+L+NL++PK+ PI GE+R AH LA+ VG+ + I++V+G+++
Sbjct: 367 GHGSARELKLMLNLMRPKPFVPIHGEFRMCHAKELAKSVGIREBAIPLVDKGGEVFRM 426

45 Query: 424 DGFFHSGVPGADVMDGNAIGDVGNIVLRKRVLSDEGPIFIVITVSKKEKKIISKARV 483
+G VP+G+V+IDG +GDVGNIVLRDR++LS+DGI +VV+T++K+ I+S +
Sbjct: 427 GQGRKAGKVPSONVLIDGLGVGDVGNIVLRDLRLSKDGLIVVVVTLINKSGOTLLSGNI 486

Query: 484 NTRGFVYVKSRLILRESAELVMTTVEYLSKDTFDMGELKGVKDEVSKFLPQCKRRP 543
+RGFVYV++S ++ E+ ELV T++ +++ +W LK VR+ +S+FL++TTRRP
Sbjct: 487 ISRGFVYVRESEKLIEANELVETLKKCVTENVMNSLSKNREVLRSRFLFKETRRP 546

Query: 544 AILPVMMEV 552
ILP++MEV
Sbjct: 547 MILPIMEV 555

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2723> which encodes the amino acid sequence <SEQ ID 2724>. Analysis of this protein sequence reveals the following:

Possible site: 33

55 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2204 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GF:BA06117 GB:AP001515 unknown conserved protein [Bacillus halodurans]
Identities = 232/549 (42%), Positives = 360/549 (65%), Gaps = 2/549 (0%)

-990-

5	Query: 4	IKMIALGGVREYKKNFYLVINDSMFILDAGLKYPENQLGVDLVINLQYVIREKKGKQ 63
	Sbjct: 9	I++ ALGGV E GN Y-VR++D +P++DAGL +P++E LGVD+VLP++ Y++E++ +V+ IKVFALGGVREYKKNFYLVINDSMFILDAGLKYPENQLGVDLVINLQYVIREKKEVR 68
10	Query: 64	GIFLSHGHADAIGALPYLLAEVSAFPVGSPLTIELAKLVKSNSTYKKNFNHVVDSDE 123
	Sbjct: 69	I L+HG D IG LPY+L +++ PV+G+LP+ L + +K + +D8++ AILLTHGHEDHIGGLPYVLQKLVNFPVYGTYLTLGLVSEKLEAGLIRSAK-LKLDSDNR 127
15	Query: 124	IEEFKGLVSPFKTTHSIPESMGIVIGTQKGNIIYTGDFKPDQAAREGYQTDLLRLASIG 183
	Sbjct: 128	++ VSFRT HSIP+S+GI I T +G I++TGDFKPDQ +G Q ++ ++A IG LKLGSSTPVSPFKTHSIPDSVGLCIQTSSQGIIVHGDGFKPDQTPVDGRQABIRGMALGH 187
20	Query: 184	EGVALLSDSVATENDQIASSEFVCEEMDSVIDADGRVIVAASNLVRIQQVDFEAT 243
	Sbjct: 188	+GVL LLSDS NA SE+EVG + GR+IV ASN+ R+QXV +A KGVLCLLSDSTNAKKRPGMTKSTEVGRGIAEAFQTKGRILVITFASNVHRVQVHAAI 247
25	Query: 244	AHGRFVVLGTDAENIVRTALREKLMITDERLLIKPKDMKSPEDHLLILEAGRWGEPT 303
	Sbjct: 248	A R++ + G +V ARL L D+ L I ++GK++D+ I+ G GRP+ ATNRKLVAGRSNVKVVSIARLGLYLAADD-LFIDIEEVSKTDDERVAIITTSQQRPM 306
30	Query: 304	NSLQIWAAGRRHYVQIKSGDLVYIVTTPSTAKRMVARVENLIYKAGGSVKLITQNLAV 363
	Sbjct: 307	++L +NA G HR + I E D V I TP E V+ +L+++ G V + S SALSRYAKGAHRQITITENVTIAATPIPGNRSVSTIVDLLHRIGADVIPGHRVNAS 366
35	Query: 364	GHANGRLQLMLNLKQYLPVFGVEYRDLAAHAKLAEEVCIPTENIHILKRGDIMVLND 423
	Sbjct: 367	GH + +L+L++NL++P++ P+ GR+R AH +LA+ VGI E I ++ +G+++ + GHGSAKLLKIMNMRPKFPVPIHGEFRMCHAKBLAKSVGIREEAFPLVDKGEVVEFRN 426
40	Query: 424	EGFLHEGGVPASDVMDNGAIDGVNIVLDRKVLSDGFIIVATVTKKKRIISKAV 483
	Sbjct: 427	G VP+ +V+IDG +GDGVNIVLDR++LS+DGI +V T++K+ I+S + GQGRKAGKVPBGVNLIDGLGVGDGVNIVLDRLLSKDGLVVVVTLNKSGTLIGENI 486
45	Query: 484	NTRGFVYVKKSHDLRESAELNITVGNLYKDDTDMGELKGNVREDDSKFLFEQTKRRP 543
	Sbjct: 487	+RGFVTV++S ++ E+ ELV T+ ++ +W LK NVR+ LS+FLFE+TKRRP ISRGFVTVRESEKLIEANELVETLTKKCTTENVNENMSLEKSNREVLGRFLFEKTKRRP 546
50	Query: 544	AILEFVMEV 552
	Sbjct: 547	ILP++MEV 555

An alignment of the GAS and GBS proteins is shown below.

Identities = 446/553 (80%), Positives = 513/553 (92%)

45	Query: 1	MSDIKIMALGGVREYKKNFYLVINDSMFILDAGLKYPENQLGVDLVINLQYVIREKKGKQ 60
	Sbjct: 1	M+DIK++ALGGVRE GN Y+VE+ND+S+P+LDAGLKYPENQLGVD+VLPNDY+IENK 60 MTDIKIALGGVREYKKNFYLVINDSMFILDAGLKYPENQLGVDLVINLQYVIREKKGKQ 60
50	Query: 61	RVGGIFLTHGHADAIGALEFYIAEVGAFPVGSPLTIELAKLVKSNSTYKKNFNHVIDS 120
	Sbjct: 61	+VGGIFL+HGHAADAIGALEFY++AEV APVGS LTIELAKLVK++ + KKNFNHVA+DS KVGGIFLSHGHADAIGALPYLLAEVSAFPVGSPLTIELAKLVKSNSTYKKNFNHVIDS 120
55	Query: 121	ETIEFPQDAVISPFKTHSIPESMGIVIGTQKGNIIYTGDFKPDQAARKYYQTDLRLAE 180
	Sbjct: 121	+TIEFPD ++SFP+TTHSIPESMGIVIGT +GNI+YTGDFKPDQAAR+ YQTDL RLAE DTIEFPKGLVSPFKTTHSIPESMGIVIGTQKGNIIYTGDFKPDQAAREGYQTDLRLAE 180
60	Query: 181	IGRDGVLALLSDSNATSNBQVASEYVEGDEIKSVIREDAGGRVIVAASNLIRIQQVDF 240
	Sbjct: 181	IG+GVLALLSDS NATSN+Q+ASE EVG+E+ SVI DA+GRVIVAASNL+RIQQVDF IQKEGVLALLSDSNATSNQIASSEFVCEEMDSVIDADGRVIVAASNLVRIQQVDF 240
65	Query: 241	AAARNGRRVVLGTDFDIENIVRTAIRMKRIADENMLIKPKDMTRYKDELILETGRMG 300
	Sbjct: 241	+A +GRRVVLGT D ENIVRTA+R++++ I DE ++IKPKDM++ED+IL+ELE GRMG SATAHGRVVLGTDAENIVRTALREKLMITDERLLIKPKDMKSPEDHLLILEAGRWGEPT 300
70	Query: 301	EPINGLQKMAIGRRHYVQIKGDLVYIVTTPSIAKEAVVARVENLIYKAGGSVKLITQNL 360
	Sbjct: 301	EPIN LQKMA GRRHYVQIK+GDLV+IVTTPS AKRA+VARVENLIYKAGGSVKLITQNL EPINSLQKMAIGRRHYVQIKGDLVYIVTTPSTAKRMVARVENLIYKAGGSVKLITQNL 360

Query: 361 RVSGHANGRELQLLMNLLKPKYLPFPQGEYRDLGAHAGLAQEVGMSADDIYIVKRGDIMV 420
 RVSGHANGRL+LQLLMNLLKFP+YLPFP+QGEYRDL+AHA LA+EWG+ ++I+I+KRGDIMV
 Sbjct: 361 RVSGHANGRLQLLMNLLKPKYLPFPQGEYRDLAAHAKLAKEVGIIPPNTIHLKRGDIMV 420

Query: 421 LEKDGPFFHSGSVPAQDWMIDGNAIGDVGNIIVLRDRKVLSEDDGIFIVITVSKKREKIISK 480
 L +GF H G VPA DWMIDGNAIGDVGNIIVLRDRKVLSEDDGIFIVITVSKKREKIISK
 Sbjct: 421 LNDEGFLHEGCVASDWMIDGNAIGDVGNIIVLRDRKVLSEDDGIFIVITVSKKREKIISK 480

Query: 481 ARVNTGRFVYVKKSRDILRESAELVNTTVEDYLSKDPFDWGELKSKVRDEVKFKFLDQTK 540
 A+VNTGRFVYVKKSRDILRESAELVNTTV+YL KDPFDWGELKSG VRD++SKFLF+QTK
 Sbjct: 481 AKVNTGRFVYVKKSRDILRESAELVNTTVGNLYKKDPFDWGELKSGVRDDEKFLFQTK 540

Query: 541 RRPAILFVVMVEVR 553
 RRPAILFVVMVEVR
 Sbjct: 541 RRPAILFVVMVEVR 553

There is also homology to SEQ ID 4910.

SEQ ID 2722 (GBS295) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 48 (lane 2; MW 89.4kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 167 (lane 9 & 11; MW 79kDa – thioredoxin fusion) and in Figure 238 (lane 3; MW 79kDa – thioredoxin fusion).

Purified Thio-GBS295-His is shown in Figure 244, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 896

A DNA sequence (GBSx0950) was identified in *S. agalactiae* <SEQ ID 2725> which encodes the amino acid sequence <SEQ ID 2726>. This protein is predicted to be tributyrin esterase. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9885> which encodes amino acid sequence <SEQ ID 9886> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF62859 GB:AF157484 tributyrin esterase [Lactococcus lactis
 subsp. lactis]
 Identities = 154/262 (58%), Positives = 198/262 (70%), Gaps = 4/262 (1%)

Query: 21 MAFNIEYHSEKVLGTERQVNVVIYDPAFEMSDDKIDCDIPVLYLLHGMGNGENSWCKRTN 80
 MA NIEY+S+VLG R+RVVIY++ ++ D DIPVLYLLHGM GNGENSW R+
 Sbjct: 1 MAVINIEYSEVLGMNRKVNVIYPSGKVED--PTQTDIPVLYLLHGMGNGENSWILRS 58

Query: 81 IERLLRHTNLIVMPSTDLAWYNTTKYGLDYDPAIALKPKVLKRFPPHMSDKRENFIA 140
 IERLLRHTNL+VMPSTDL+Y NT YG++YDPAIA ELPKV+ FFPH+S KRENFIA
 Sbjct: 59 IERLLRHTNLAIVMPSTDLGFYNTTKYGLDYDPAIAELPKVINFNFHPLSTKRENFIA 118

Query: 141 GLGSGYGYAYKIALITNRFHGAALSGALSGDFDILLPNNNNNNIYWSGIPGDAINTINI 200
 GLGSGYGYAY++AL T+ FG+AALSG L+FD + N N YW GIPG+

-992-

Sbjct: 119 GLSMGGYGAYRLALGSLTDYPSYAASLSGVLTFDG--MEENFKENPAYWGGIFGNWEIFKGS 176
 Query: 201 ERHSRLRYVESFDMKTKFYAWCGYEDFLFEANEVAIDELRQLGLTIDYFNDHGKHGWYYW 260
 + L + + K K YAWCG +DPLF NR A EL++LG I Y + G HEMWYW
 Sbjct: 177 DNEILSLADRKQEMKPKLYAWCGKQDPLFPGRBYNTAEKLGFDITYESSDGVHEWYWW 236
 Query: 261 NQQLKKVLEMLPVDYVKEERLS 282
 Q++E VL+MLP++Y +EERLS
 Sbjct: 237 TQKIESVLKPLINYNKEERLS 258

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2727> which encodes the amino acid sequence <SEQ ID 2728>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2183 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 172/262 (65%), Positives = 199/262 (75%), Gaps = 1/262 (0%)
 Query: 21 MAFNIEYHSKVLGTERQNNVIYPDAFMSDDKIDDCDIPVLYLLHGMGGNENSWQKRTN 80
 NA IEYHS VLG ER+NNVIYPD E+ D DIPVLYLLHGMGGNENSWQKRT
 Sbjct: 1 MASIAIEYHSVVLGMRKKNVVIYPDQSEIPKKDQSKDIPVLYLLHGMGGNENSWQKRTA 60
 Query: 81 IERLLRHTNLIVMPSTDLAWYTTKYGLDYFDAIAELPKVLKRPFPNMSDKREKNFIA 140
 IERLLRHTNLIVMPSTDL WYTT YGL+Y+ A++ ELP+VL FFPNM+ KREK P+A
 Sbjct: 61 IERLLRHTNLIVMPSTDLGWTYDTAYGLNRYALSQELRQVLAAPFPMTQKRETFVA 120
 Query: 141 GLSMGGYGAYKIALNLNRFPSHAASLSGALSPDFDILLFNKGNNNINYSWGFGLNNTDNI 200
 GLSMGGYGAY+K AL +NRFPS+AA8 SGAL P + L + YW G+FG ++ D +
 Sbjct: 121 GLSMGGYGAFIMALKENRFPSYAASPSGALDFSPETLLEKGLGELAYWQGVFGQDFD-L 179
 Query: 201 ERHSRLRYVESFDMKTKFYAWCGYEDFLFEANEVAIDELRQLGLTIDYFNDHGKHGWYYW 260
 ++H L+ V D KTKFYAWCGYEDFLF NE AI + + GL IDY HGKHGWYYW
 Sbjct: 180 DKHYLKNMVAESDGGTKFYAWCGYEDFLFATNEKAIADFOAQGLDIDYHKGHGHEWYYW 239
 Query: 261 NQQLKKVLEMLPVDYVKEERLS 282
 NQQL +LEMLP++Y KEERLS
 Sbjct: 240 NQQLVLEMLPVDYVKEERLS 261

SEQ ID 2726 (GBS645) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 129 (lanes 8 & 10; MW 60kDa + lane 9; MW 27kDa) and in Figure 186 (lane 4; MW 60kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 129 (lane 12; MW 34.7kDa), in Figure 140 (lane 8; MW 35kDa) and in Figure 178 (lane 4; MW 35kDa). Purified GBS645-GST is shown in Figure 236, lane 11; purified GBS645-His is shown in Figure 229, lanes 3-4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 897

A DNA sequence (GBSx0951) was identified in *S.agalactiae* <SEQ ID 2729> which encodes the amino acid sequence <SEQ ID 2730>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

-993-

```

INTEGRAL    Likelihood = -9.34    Transmembrane    22 - 38 ( 18 - 46)

----- Final Results -----
bacterial membrane --- Certainty=0.4736 (Affirmative) < succ>
5 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2731> which encodes the amino acid sequence <SEQ ID 2732>. Analysis of this protein sequence reveals the following:

```

Possible site: 52

>>> Seems to have an uncleavable N-term signal seq
INTEGRAL    Likelihood = -7.43    Transmembrane    25 - 41 ( 20 - 46)
15 INTEGRAL    Likelihood = -2.71    Transmembrane    4 - 20 ( 3 - 20)

----- Final Results -----
bacterial membrane --- Certainty=0.3972 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
20 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GST proteins is shown below.

```

Identities = 31/87 (35%), Positives = 50/87 (56%), Gaps = 2/87 (2%)

25 Query: 1 MRTLRFMIFALPKFIFRLIWNIIIGFKTVLVIAILFLGLYYVYANHSQSEFANQLSDIIQ 60
      M+ L +I +PK I ++ W++I G +T+L++ II+ GL YY+NHS S AN++S I
Sbjct: 1 MKQLLAIIWLPLKLVIVMPFWHLIKGLFQITILLVTIIIGIMTYSNHSDSVLANKIS--IV 58

30 Query: 61 TGTFTLNFDATNQLQNSFTNLATDNVH 87
      T + F Q ++ T + N H
Sbjct: 59 TEQVVCIFDILTQKPSAKTRHSGGNSH 85

```

SEQ ID 2730 (GBS220d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 11-13; MW 50kDa) and in Figure 239 (lane 12; MW 50kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 14-16; MW 25.2kDa) and in Figure 184 (lane 7; MW 25kDa). Purified GBS220d-GST is shown in Figure 246, lanes 3 & 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 898

A DNA sequence (GBSx0953) was identified in *S.agalactiae* <SEQ ID 2733> which encodes the amino acid sequence <SEQ ID 2734>. This protein is predicted to be unnamed protein product (rpiA). Analysis of this protein sequence reveals the following:

```

45 Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2538 (Affirmative) < succ>
50 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

-994-

>GP:CA869583 GB:A93589 unnamed protein product [Spinacia oleracea]
 Identities = 114/232 (49%), Positives = 147/232 (63%), Gaps = 11/232 (4%)

Query: 2 DELKKLAGVTAQYVKNKMGIVGLGTGSTAYFFVEIGRRVKEBGLQVVGVTTSNRTTEQ 60
 D+LKKLA A VK+GM+GLGTGSTA F V IG + L +VG+ TS RT BQ
 Sbjct: 59 DDLKKLAABKAVDSVKSQVGLGTGSTAFAVSRIGELLSAGKLTNIVGIIPTSKRTARQ 118

Query: 61 ARGGLGIPLKSADDIDVIDVTVDGADEVDPDFNGIKGGGALLMEKIVATPTKEVINVVDR 120
 A LGIPL DD ID+ +DGADEVDPD N +KG GGALL EK+V + +I VVD+
 Sbjct: 119 AASLGIPLSVLDHPRIDLADGADEVDPDLNIVKRGGALLREKVMREASDKRTIVVDD 178

Query: 121 SKLVEITLGAPFL--PVEVV---RYGSERLFRVFKSGYCPSPRETEGDR--FITMGNY 172
 +KLV+ LG +L PVEVV +Y +RL +FK G C + BGD ++TD NY
 Sbjct: 179 TKLVDLGLGSGRIAMPVEVVQPCWKYNLKRLOEIKRLG-CRAKLRMBGDSPPYITNSNY 237

Query: 173 IIDLKL-KKIEDPKQLANELDITGVVVEHGLFNGMKNKIVAGKNGLDILEK 223
 I+DL I+D + E+ GVVEHGLF GM ++VI+AGK G+ + K
 Sbjct: 238 IVDLYFFPTSIKDARAAGREISALEBVGVEHGLFLGMASEVITLAKTGVSVKTK 289

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2735> which encodes the amino acid sequence <SEQ ID 2736>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1646 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 166/222 (74%), Positives = 190/222 (84%)

Query: 1 MDLKKLAGVTAQYVKNKMGIVGLGTGSTAYFFVEIGRRVKEBGLQVVGVTTSNRTTEQ 60
 N+ LKX+AGVITAA+YV +GM +GLGTGSTAY+FVEIGRRVK+BGLQVVGVTTS+ T+Q
 Sbjct: 1 MEALKKLAGVTAQYVITDMITGLGTGSTAYFVEIGRRVKGBGLQVVGVTTSVTSKQ 60

Query: 61 ARGGLGIPLKSADDIDVIDVTVDGADEVDPDFNGIKGGGALLMEKIVATPTKEVINVVDR 120
 A LGIPLKS DD ID+TVDGADEVDPDFNGIKGGGALLMEKIVATPTKEVINVVDR 120
 Sbjct: 61 AAVLGIPLKSIDDIDSIDLTVDGADEVDPDFNGIKGGGALLMEKIVATPTKEVINVVDR 120

Query: 121 SKLVEITLGAPFLPVEVVRYGSERLFRVFKSGYCPSPRETEGDRFITMGNYIIDLKK 180
 SK+VE LGAPFLPVEVV+YG+RLFRV+ GY PSFR R +TDM NYIIDLKL
 Sbjct: 121 SKMVEHLLGAPFLPVEVVQYADRLFRVFKSGYKPSFRMGKDSRLVITDMQNYIIDLGLQ 180

Query: 181 IEDPKQLANELDITGVVVEHGLFNGMKNKIVAGKNGLDILE 222
 I+DP + LD TVGVVEHGLFNGM+KVIVA K+G+ +LE
 Sbjct: 181 IKDPVAPGHLLDITGVVVEHGLFNGMVDKIVASDGVTVLE 222

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 899

A DNA sequence (GBSx0954) was identified in *S.agalactiae* <SEQ ID 2737> which encodes the amino acid sequence <SEQ ID 2738>. This protein is predicted to be phosphopentomutase (deoB). Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0546 (Affirmative) < succ>

-995-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AA045496 GB:U00410 phosphopentomutase [*Lactococcus lactis* subsp. cremoris]
 Identities = 275/408 (67%), Positives = 325/408 (79%), Gaps = 7/408 (1%)

Query: 3 QFDRHLVVLDSVGIGAAPDANDFVNAGVP-----DGASDTLGHISKTIVGLAVPMNAKI 56
 +F RIHLAV+DSVGIGAAPDA+ F N V D SDT+GHIS+ GL VPM+ K+
 10 Sbjct: 4 KFGRIHLVMDSDVGIGAAPADKFPNHDVETHRANDVKSDTIGHISEIRGLDVPNTKLI 63

Query: 57 GLGNIPRQALKTVPAENPESGYATKLQEVSLGKDTMTGHWIEMGLNITEPDTFVNGFF 116
 G GNIPR LKT+PA + P+ Y TKL+E+S GKDITGHWIEMGLNI PF T+ G+P
 15 Sbjct: 64 GWNIPRPSPLKTIPTAAQKPAAYVTKLEIRISGKDTMTGHWIEMGLNITPFPPTVPEGT 123

Query: 117 EDITIKIEDFSGRKVIREANKPYSGTAVIDDFGPRQMETGELI IYTSADFVLQIAAHEDI 176
 ED++ KIE+PSGRK+IREANKPYSGTAVI+DFGPRQ+ETGELI IYTSADFVLQIAAHED+
 20 Sbjct: 124 EDLLEKIEEFSGRKIIIREANKPYSGTAVIEDFGPRQLETGELI IYTSADFVLQIAAHEDV 183

Query: 177 IPLEELYRICYARSITMERFALL-GRIIARPYVGEFGNFTKTANRHDYAVSPFEDTVLN 235
 I EELY+ICEY RSIT+E ++ GRIIARPYVGE GNF RT R DYA+SPF +TVL
 25 Sbjct: 184 ISREELYKICEYRSITLESIGMIGRIIARPYVGEAGNFERTDGRDVALSPFAPTVL 243

Query: 236 KLDQAGIDITYAVGKINDIFNGSGINHDGHNKNSHGIDTLTKTMGLSEFEKGFSPTNLV 295
 KL +AGIDTY+VGKI+DIFN G+ +DMGHN ++ G+D L+K M +EF +GFSPTNLV
 25 Sbjct: 244 KLYKAGIDITYVGKISIDIPNTGVKYGWGHNDNDGVDRLKAMKTFTBGSFPTNLV 303

Query: 296 DFDALYGHRRDPHGYRDLHEPDERLPEIISAMRDKDLLLTADHGNQDPTAGTQHTREY 355
 DFD A YGHRRD GY + +FD RLPEII AM++ DLL-ITADHGNQD+Y GTDHTREY
 30 Sbjct: 304 DFDALYGHRRDVEGYKALIEDPGRGLPEIISAMRDKDLLMTADHGNQDPSYVGTQHTREY 363

Query: 356 IELLAYSPTSFTNGILIPVGHFADISATVADNFQVDTAMIGESFLQDLV 403
 IEL+ +S SF ++PVGHFADISAT+A+NF V A GESFL LV
 35 Sbjct: 364 IELVIPSFSFKEPKVLVPGHFADISATIAENPFSVTKAQTGESFLDALV 411

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2739> which encodes the amino acid sequence <SEQ ID 2740>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0185 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 348/402 (86%), Positives = 374/402 (92%)

50 Query: 1 MSQFDRHLVVLDSVGIGAAPDANDFVNAGVPDASDTLGHISKTIVGLAVPMNAKIGLGN 60
 MS+F+RIHLVVLDSVGIGAAPDA+ F NAGV D SDT+LGHIS+ GL+VPMNAKIGLGN
 Sbjct: 1 MSKFNRIHLVVLDSVGIGAAPADKFPNHDVETHRANDVKSDTIGHISEIRGLDVPNTKLI 60

Query: 61 IPRPQALKTVPAENPESGYATKLQEVSLGKDTMTGHWIEMGLNITEPDTFVNGFFEDI 120
 I RP LKTVP E+NP+GY TKL+EVSLGKDTMTGHWIEMGLNITEPDTFVNGFFPE+I+
 55 Sbjct: 61 ISRPILKTVPTEDNPTGYVTKLEIRISGKDTMTGHWIEMGLNITSPFTFVNGFFPEIL 120

Query: 121 TKIEDFSGRKVIREANKPYSGTAVIDDFGPRQMETGELI IYTSADFVLQIAAHEDIIPLE 180
 TKIE+FSGRK+IREANKPYSGTAVIDDFGPRQMETGELI+YTSADFVLQIAAHEDIIP+E
 60 Sbjct: 121 TKIEFSGRKIIIREANKPYSGTAVIDDFGPRQMETGELI IYTSADFVLQIAAHEDIIPVE 180

Query: 181 ELYRICYARSITMERFALLGRIIARPYVGEFGNFTKTANRHDYAVSPFEDTVLNKLOQA 240
 ELY+I+CYARSIT+ERFALLGRIIARPYVGEFGNFTKTANRHDYAVSPF+DTVLNKL A
 Sbjct: 181 ELYKICEYARSITLERFALLGRIIARPYVGEFGNFTKTANRHDYAVSPFQDTVLNKLADA 240

-996-

Query: 241 GIDTVAVGKINDIFPKSGSINHDGHNKSNHGIDTLLIKTGLSEFEKGFSTNLVDPIAL 300
 G+ TYAVGKINDIFPKSGI +DMGHNKSNHGIDTLLIKT+ L RF KGPSFTNLVDPIA
 Sbjct: 241 GVPTYAVGKINDIFPKSGITNDMGHNKSNHGIDTLLIKTLQLPEFTKGPSFTNLVDPIAN 300

5 Query: 301 YGHRDPHGVRDCLHEFDRLPETISAMRDKILLITADHGNDPTYAGTDHTRYIPLLA 360
 +GHRDPD GYRDCLEHFD RLPRII+ M++ DILLITADHGNDPTYAGTDHTRYIPLLA
 Sbjct: 301 PGHRDPGEGYRDCLEHFDNRLPRIIANMKEDILLITADHGNDPTYAGTDHTRYIPLLA 360

10 Query: 361 YSPSTQNGILPVGHFADISATVARNPGVDTAMIGSFQLDL 402
 YS SPTQGLIP GHFADISATV+NPFGVDTAMIGESFL L
 Sbjct: 361 YSVSPTQNGILPQGHFADISATVARNPGVDTAMIGSFSL 402

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 900

A DNA sequence (GBSx0955) was identified in *S. agalactiae* <SEQ ID 2741> which encodes the amino acid sequence <SEQ ID 2742>. This protein is predicted to be unnamed protein product (mtaP). Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.44 Transmembrane 215 - 231 (215 - 231)

----- Final Results -----
 bacterial membrane --- Certainty=0.1574 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

25 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2743> which encodes the amino acid sequence <SEQ ID 2744>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.44 Transmembrane 215 - 231 (215 - 231)

----- Final Results -----
 bacterial membrane --- Certainty=0.1574 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

40 An alignment of the GAS and GBS proteins is shown below.

Identities = 225/269 (83%), Positives = 248/269 (91%)

Query: 1 MTLEKINETKDFLQAGVTAPEFGILLGSGLGELAEETNPVVDYADIPNWOQSTVVG 60
 M+L+ KINET+DPL KG+ PRFOLLGSGLGELAE+EN IV+DYADIPNWG+STVVG
 45 Sbjct: 1 MSIMTKINETKDFLVTGKLETFEFGILLGSGLGELAEVENATVIDYADIPNWKSTVVG 60

Query: 61 HAGKLVYGDLSGRKVLALQGRFHYEGNTEVVTFFVRIMRALACHSVLVTNAAGSIGVG 120
 HAGKLVYGDLSGRKVLALQGRFHYEGN +EVVTFFVR+M+AL C VLVTVNAAGSIGVG
 Sbjct: 61 HAGKLVYGDLAGRKVLALQGRFHYEGNPLEVVTFFVRVMKALGCKRSVLVTNAAGSIGVG 120

50 Query: 121 PGTLMLKDHINMIGTNPLIGENLEBPGRPFDMSDAYTATYRQKAHQIAEKQNIKLEEG 180
 PGTLI DHIM G NPLIGENL+EPGPRFPDMSDAYT YR KAH++ARK NIKLE+G
 Sbjct: 121 PGTLGATIDHINMTGNPLIGENLDKPGRPFDMSDAYTKVYRNKQHEVAEKQNIKLEEG 180

55 Query: 181 VYLGVSQFTYETPAEIRAFQIMGAQAVGMSTVPEVIVAHSGLKVLGISAITNPAAGFQS 240
 VY+G+G+GFTYETPAEIRAF+ +GA AVGMSTVPEVIVAHSGLKVLGISAITNPAAGFQS
 Sbjct: 181 VYMLTGFTYETPAEIRAFKVLGADAVGMSTVPEVIVAHSGLKVLGISAITNPAAGFQS 240

Query: 241 ELNHEEVETQRIKEDPKGLVKSIVARL 269

-997-

ELNHEEVVEVTQ IKEDFKGLVK++ARL
 Sbjct: 241 ELNHEEVVEVTQHIEDFKGLVKAILAEL 269

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 901

A DNA sequence (GBSx0956) was identified in *S. agalactiae* <SEQ ID 2745> which encodes the amino acid sequence <SEQ ID 2746>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have a cleavable N-term signal seq.
 10 INTEGRAL Likelihood = -9.34 Transmembrane 266 - 282 (263 - 289)
 INTEGRAL Likelihood = -8.97 Transmembrane 231 - 247 (229 - 253)
 INTEGRAL Likelihood = -7.70 Transmembrane 356 - 372 (352 - 376)
 15 INTEGRAL Likelihood = -7.32 Transmembrane 303 - 319 (297 - 326)
 INTEGRAL Likelihood = -5.57 Transmembrane 337 - 353 (334 - 355)
 INTEGRAL Likelihood = -5.57 Transmembrane 391 - 407 (387 - 409)
 INTEGRAL Likelihood = -2.44 Transmembrane 177 - 193 (177 - 193)
 INTEGRAL Likelihood = -1.01 Transmembrane 159 - 175 (159 - 175)
 20 INTEGRAL Likelihood = -0.43 Transmembrane 198 - 214 (196 - 215)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4736(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9883> which encodes amino acid sequence <SEQ ID 9884> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP: AAD53928 (B:AF179611 chloride channel protein [Zymomonas
 30 mobilis]
 Identities = 121/410 (29%), Positives = 213/410 (51%), Gaps = 19/410 (4%)
 Query: 14 VKFMIAVFMVWAGVGAIIAHVVLAPTEWIAFGDSRENTGLIN-----SVTPIKRVL 67
 +++ +A L + + G+G +L+ +L + +H+G S ++ +S + + +P+R+
 35 Sbjct: 3 IRYGLACLVAGCLTGIGMLLSWILHVAQHIAHYGSLQHVISESFLKGSAAFPRLLE 62
 Query: 68 SLTLVSLFASLWVYIQLPKQITSIKQVVFKDPVKSPLYMLHGHAFQLQLIVGTG 127
 L + W L + + SI Q V + P+W I H LQ+ + V G G
 40 Sbjct: 63 VLVFCRAVVGCGWGLRHFGSPLVSIQAVANK---RVMPFNTIIHVLLQITVVLGS 119
 Query: 128 PTGKEGAFREPGAINAGKISDLALKVLDRILLISGAAGLGAVVQVPLASVPFAFET 187
 P+G+E APRE G++ + + L +R+L+ GA AG ++VY VPL+ FA E L
 Sbjct: 120 PLGRVAPRHLGLIGERFAPWGLSENQRILVACGAGAGFASVYNNVPLSGALFALEAL 179
 45 Query: 188 ALGISLKNITVTHASTFCASTIAQLVISTAPLVHISIMSLNSQSIAFMPLIVLCVPTI-- 245
 + + + + L ++ +A +A +++ +YH+ ++++ +L+ L PI
 Sbjct: 180 LMTWASPVVIVALLTSALSARMWILLGNSMVVHVHPAMPVDTR---LMLLALLAQPIFG 235
 Query: 246 --AISFYLNQKVTERRIK-NIKILLSLPVVSLIVSLVSIYVQILGKGLALWQEVFGT 302
 A FR+ +QK+T RIK N++ L + + +LS+ +P+ILGNG V F
 50 Sbjct: 236 IAAHYFRFWQKITAGRIKRNRIALVAILCFAAIGLSMWFFELONGKGPVSLAFEN 295
 Query: 303 TVSLIA-IILVIAIMIAITSLTYAGYGGILTPFSFISGACLFGLASISIPALPHISVTS 361
 + A L K++A I-AGAYGG+LTP S GA L ++ + LP + I
 55 Sbjct: 296 LSGMKAGELFCPKILAVFLAKWAGYGLLTIPGISFCALLAVVIGHLHNMWLPVPTIGAF 355
 Query: 362 MIVGAALFLATMRAPLTAUGLVISPIQGSVITIVPLTIAVLATAYDYDF 411
 ++G A FLA +M+ P+TA+ LVI F ++P+ AV +A F
 Sbjct: 356 AIIIGAAPLASKMPITMMLVIEPARTGIDFLPIAFVAGSIAISQF 405

60

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2747> which encodes the amino acid sequence <SEQ ID 2748>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence
5  INTEGRAL    Likelihood = -5.41    Transmembrane    247 - 263 ( 245 - 267)
   INTEGRAL    Likelihood = -5.15    Transmembrane    326 - 342 ( 323 - 345)
   INTEGRAL    Likelihood = -5.04    Transmembrane    411 - 427 ( 407 - 429)
   INTEGRAL    Likelihood = -4.94    Transmembrane    39 - 55 ( 34 - 59)
10  INTEGRAL    Likelihood = -4.46    Transmembrane    284 - 300 ( 282 - 307)
   INTEGRAL    Likelihood = -3.45    Transmembrane    380 - 396 ( 376 - 400)
   INTEGRAL    Likelihood = -2.13    Transmembrane    185 - 201 ( 184 - 201)
   INTEGRAL    Likelihood = -2.02    Transmembrane    88 - 104 ( 87 - 105)
   INTEGRAL    Likelihood = -1.12    Transmembrane    350 - 366 ( 350 - 367)

15  ----- Final Results -----
      bacterial membrane --- Certainty=0.3166 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

20  The protein has homology with the following sequences in the databases:
>GP:AAF41386 GB:AE002449 chloride channel protein-related protein
[Neisseria meningitidis MC58]
Identities = 137/373 (36%), Positives = 201/373 (53%), Gaps = 23/373 (6%)

25  Query: 59  IHLIQSLSPFGSQG---SFSIMIASVPPORRALSLFAGLLAGLQWHLAKKKGDIQSI 114
      +H IQ ++G+ SF +A RR L G +AG GW LL + GK I
Sbjct: 1'  MHFIQHTAYGYGADGVYTSFRBVGQASGMRRFVALTLOGAVAGSGWMLKRFKPKQIEI 60

Query: 115  QQIIQDDISFSFSPW-TQFWHGMQLQTTVMGAPVREGASREAVAVTISLMSQRCNLKAD 173
      + ++ + P+ T +H LQ+ TV +G+P+GRE A RE+ +R L + +
30  Sbjct: 61  KAALKQLQLQPLFLITVHFVHLQITVGLQSLPREGVAPREMTNAFAFGGKRLGLDGE 120

Query: 174  QKILLACASGAAGAVYNAPLATILFILEAILNRMSLKNIYAACLTISYVAVETVALLQGR 233
      +LL+ACASGA L AVYN PLA+ LFILEA+L W+ + + AA LTS +A + G
35  Sbjct: 121  MRILLACASGAGLAAYVYNPLASTLFILEAMGLQVVTQQAVALITSVIATAVARI--GL 178

Query: 234  HEIQYLMPOQHWTLTGT--LIGSVLAGLILSLFAHAYKHLLKHLPKADAKSQNPFPKVLIA 291
      ++Q P + T+ T L S + G IL + A ++ + P + IP +
40  Sbjct: 179  GDVQGYHP-ANLTVNTSLMFSAVIGPILGVAAVFPQRTAQKPFPIKRDNIKIPLAVCN 237

Query: 292  FSLIAGLSIFPPEILGNKGAG--LLP-FLHEEPH---LSYISWLINAKAVASLNFASGA 345
      F+LI +S++FPEILGNKGAG L P L + H L+ +WL+V A+A+ GA
45  Sbjct: 238  FALIGVISVWFPEILGNKGAGNQLTGGSLTDWQSLGLTAVKVLVNLMAV-----GA 291

Query: 346  KGCKIAPSNMGGASGLLLAILSQYLIPLSLNTLAIMVGATITPLGVINKIPLAAPVFLV 405
      GG I PNMILG A + P +S+ A +VGA +PLGV K+PL A P++
50  Sbjct: 292  YGGLITPSNMGLSTIAFAAATAMISVFP-EMSSESAIVGAAVFLGVSLKMLTALAFIL 350

Query: 406  EITGQSLLMIIPL 418
      E+T + +++PL
50  Sbjct: 351  ELTYAPVALMIPL 363

Identities = 131/415 (31%), Positives = 215/415 (51%), Gaps = 9/415 (2%)

55  Query: 2  LNFKMVSRLYYAVKFMIAVLFWMT-VMGVGAILMHYVLMFTEPLAGDSRENTISLLISV 60
      LNF S + + LF+T + AG+ A ++ + + L+FG S+ + +++ SV
Sbjct: 22  LNFCSNLSMKKHFLILTFYLFILGLTAGLVAFILTKAHLIQSLSPFGSQSFSFIMIASV 81

60  Query: 61  TPIKRVLSLITWSFLASISWYLLQKFKQITKISIQQVVKQPSVKKSPYWHHGHAFIQL 120
      P +R LSL LA L W+ L K K I SI QQ++ D S SP W H +LQL
Sbjct: 82  PPQRRALSILFAGLLAGLQWHLAKKKGDIQSI-QQIIQDDISF--SP-WTQFWHGLQL 137

Query: 121  IYVGTGSPIGKGAAPREGPAINKKSIDLLALKVLDRKLLISGAAAGLSAVYQVPLASV 180

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 131/415 (31%), Positives = 215/415 (51%), Gaps = 9/415 (2%)

55  Query: 2  LNFKMVSRLYYAVKFMIAVLFWMT-VMGVGAILMHYVLMFTEPLAGDSRENTISLLISV 60
      LNF S + + LF+T + AG+ A ++ + + L+FG S+ + +++ SV
Sbjct: 22  LNFCSNLSMKKHFLILTFYLFILGLTAGLVAFILTKAHLIQSLSPFGSQSFSFIMIASV 81

60  Query: 61  TPIKRVLSLITWSFLASISWYLLQKFKQITKISIQQVVKQPSVKKSPYWHHGHAFIQL 120
      P +R LSL LA L W+ L K K I SI QQ++ D S SP W H +LQL
Sbjct: 82  PPQRRALSILFAGLLAGLQWHLAKKKGDIQSI-QQIIQDDISF--SP-WTQFWHGLQL 137

Query: 121  IYVGTGSPIGKGAAPREGPAINKKSIDLLALKVLDRKLLISGAAAGLSAVYQVPLASV 180

```


V G P+G+RGA RE S L D++LL+ + A L AYY PLA++
 Sbjct: 138 TTVSMGAPGVREGASREVAVTITLSPGSSQRCLNKADQKLLACAGGAALGAVYNAFLATI 197
 5 Query: 181 FFAFETLLALGISLKNITVILLASTPFGASIVLQVISTAPL-THYSKMSLMSQSLAFMFLIV 239
 P E + S+KNI +++ A V T++L + Y + + ++ +L L
 Sbjct: 198 LPLEIAINRNSLKNITVACLTSTVAVTALQGRHEIQYILMPQHWHTLSTLIGSVLGA 257
 Query: 240 LCVTFPIAISRYLNQKVTERRINIKILLSLSPVLSLIVSTVYPGLNLGNGA-LWQEV 298
 L ++ A +++++L ++ K+ + + + + ++ LSI + +ILNG A L + +
 10 Sbjct: 258 LILSLFAHAYKHLKLLHPKADAKS NF + PKVLIAFSLAGISIPFPBILGNCRAGILPFL 317
 Query: 299 FKGITVSLIALIIVLKMIAITLSTLYAGAYGGLITPSPSIGACGLPGLIASISIPLLP-HIS 357
 + +S I+ L+V K +A +GA GG + PS +G G LIA +S I+P +S
 Sbjct: 318 HEEPHLSYIISLWAKVAISLVSFASGAKGGLIAPSMWGASGILIAISLQYLPLSL 377
 15 Query: 358 IVTSMLVGAAIFLAIMRAPLTAVGLVISFSGQSVITITVPLTIA-VLPATADYDF 411
 +++VGA IFL + + PL A ++ TGQSG+ I+PL A++F +Y ++
 Sbjct: 378 NTLAIMVGATITPLGVINKIPLAAPFLVETVQSGLLIMTILALAMLLIPYFSQYF 422

20 A related GBS gene <SEQ ID 8683> and protein <SEQ ID 8684> were also identified. Analysis of this protein sequence reveals the following:

```

Lipp: Possible site: -1      Crend: 9
SRCFLIG: 0
McG: Length of UR: 19
25 Peak Value of UR: 2.96
    Net Charge of CR: 2
McG: Discrim Score: 9.64
GVH: Signal Score (-7.5): 1.15
    Possible site: 26
30 >>> Seems to have a cleavable N-term signal seq.
    Amino Acid Composition: calculated from 27
ALOM program count: 9 value: -9.34 threshold: 0.0
    INTEGRAL Likelihood = -9.34 Transmembrane 261 - 277 (258 - 284)
    INTEGRAL Likelihood = -8.97 Transmembrane 226 - 242 (224 - 248)
35    INTEGRAL Likelihood = -7.70 Transmembrane 351 - 367 (347 - 371)
    INTEGRAL Likelihood = -7.32 Transmembrane 298 - 314 (292 - 321)
    INTEGRAL Likelihood = -5.57 Transmembrane 332 - 348 (329 - 350)
    INTEGRAL Likelihood = -5.57 Transmembrane 366 - 402 (382 - 404)
    INTEGRAL Likelihood = -2.44 Transmembrane 172 - 188 (172 - 188)
40    INTEGRAL Likelihood = -1.01 Transmembrane 154 - 170 (154 - 170)
    INTEGRAL Likelihood = -0.43 Transmembrane 193 - 209 (191 - 210)
    PERIPHERAL Likelihood = 1.22 61
    modified ALOM score: 2.37
icml HYPID: 7 CFP: 0.474

45 *** Reasoning Step: 3

----- Final Results -----
    bacterial membrane --- Certainty=0.4736 (Affirmative) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

55 ORF00327(340 - 1533 of 1869)
GP|5834362|gb|A053928.1|AFI79611_12|AFI79611(3 - 405 of 425) %Match = 14.7
{Zymomonas mobilis}
%Identity = 30.2 %Similarity = 56.1
Matches = 121 Mismatches = 169 Conservative Sub.s = 104

60

270 300 330 360 390 420 450 468
RSLKLLSVLKKISRD*LNH*LNFNMVSRILYYAVKFMIALVMFMVAGVAILMHYIMFTENWLAFGDSRENTLS-----L
::: | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
MKIRYLGLACLAGCLTGLGMLLSWLHVAOHRJLYGYLOLVHIGESFL

65 10 20 30 40

-1001-

Query: 181 EAAALYLLAAQHVNALAMPTISDNLANPEEDTSAEERQTTFTDMMKVGLLETLISE 236
 EAAALYLLAAQH V+ALA+NTISD+L NP+EDT+AEERQ TTFTDMMKVGLLETLI++
 Sbjet: 181 EAAALYLLAAQHVDALAIHTISDGLVNPDEDTAEEERQTTFTDMMKVGLLETLIAD 236

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2751> which encodes the amino acid sequence <SEQ ID 2752>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2117(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 210/235 (89%), Positives = 226/235 (95%)

Query: 1 MSINI EAKQGEIADKILLPGDPLRAKFAIENFLEDVACFNTVRNMPGYTGTGYKHGRVSV 60
 MSINI AK+G+IADKILLPGDPLRAKFAIENFLEDVACFNTVRNMPGYTGTGYKHGRVSV 60
 Sbjet: 1 MSINI SAKGDIADKILLPGDPLRAKFAIENFLEDVACFNEVRNMPGYTGTGYKHGRVSV 60
 Query: 61 GTGNGMPISISYARELIVDYGKTLIRVGTAGAINPDHVRRLVLAQAANTNSNIIRND 120
 GTGNGMPISISYARELIVDYGKTLIRVGTAGAI+P+HVRRLVLAQAANTNSNIIRND+ 120
 Sbjet: 61 GTGNGMPISISYARELIVDYGKTLIRVGTAGAI DPEVHVRRLVLAQAANTNSNIIRND 120
 Query: 121 PEPDFPQIADPGLLDKAYNIAKEMDITTHVGSVSSDVFYSNQPDHMLGKLGVAIEM 180
 PEPDFPQIADP LLDKAYNIA+EM +TTHVG+VLSDDVFY+N P+RMLGKLGVAIEM 180
 Sbjet: 121 PEPDFPQIADPGLLDKAYNIAKEMGVTHVGVNLSDDVFYTHMPERHMLGKLGVAIEM 180
 Query: 181 EAAALYLLAAQHVNALAMPTISDNLANPEEDTSAEERQTTFTDMMKVGLLETLIS 235
 EAAALYLLAAQH+V AL +MTISDN+P EDT+AEERQTTFTDMMKVGLLETLI+ 235
 Sbjet: 181 EAAALYLLAAQHVKALGIMTISDNLANPDEDTAEEERQTTFTDMMKVGLLETLIA 235

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 903

A DNA sequence (GBSx0958) was identified in *S.agalactiae* <SEQ ID 2753> which encodes the amino acid sequence <SEQ ID 2754>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1710(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9881> which encodes amino acid sequence <SEQ ID 9882> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2755> which encodes the amino acid sequence <SEQ ID 2756>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence

-1002-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1386(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 126/253 (49%), Positives = 175/253 (68%), Gaps = 2/253 (0%)

Query: 3 IEKTDFFSTALKVLNDQSYSHAFLLQKHGFLMSDLFLLEMMKKRRRLINIDFLFAHQEQ 62
 + MT+ T L +LD Y+Y+AF + + + L+LLEMMKKRRRLIN+ FL H +
 Sbjct: 1 LPWTNNQT-LDILLDVYAYNHAFRIAKALENIPKTALEYLEMLKRRRLINLAFLRHAEE 59

Query: 63 VVILQSKYNIKL-LHNPIDLELLANYIMDLRAKVNGLIIDFVRSVSPILYRLFMILAAQ 121
 +++Y+ L L+ + E +ANYI+DLE KVNG IIDFVRSVSPILYRLF+ L+
 Sbjct: 60 NRTIEDQYHCSLWLNQSLSDSQIANYIIDLSEVKVNGIAIIDFVRSVSPILYRLFLRLITS 119

Query: 122 EYVPLHMDYIHNARDQHYDTWFKFKELKESNHPVLLAFSEWHDSRLTSKSLAECLQITDLD 181
 E+P+ YI + +D YDTW F+ + ES+H V R+ + +T+KSLA+ L LT L
 Sbjct: 120 SIPNFKAYIFDTMDQYDTWHFQAMLESDHEVFAYLSQKQSRNVTTYSLADMILTSLP 179

Query: 182 EEVKGSTIIQLRQFEKSVRNELAHLIKPFDEQEELYRTTFQSSQAFLDQIIFLAKVIGVEYD 241
 +E+K + LR FEK+VENPLAHLIKPFDE+EL+RTT FSSQAFL+ II IA GV Y
 Sbjct: 180 QEIKDLVFLLRHFEKAVENPLAHLIKPFDEEELHRTTFPSSQAFLNIITITATPSGVIIYR 239

Query: 242 TVMFHYDTVNKLI 254
 F+D +N +I
 Sbjct: 240 REPFYDDMNAILI 252

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 904

A DNA sequence (GBSx0959) was identified in *S.agalactiae* <SEQ ID 2757> which encodes the amino acid sequence <SEQ ID 2758>. This protein is predicted to be CpsY protein. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.59 Transmembrane 260 - 276 (260 - 276)

----- Final Results -----

bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9879> which encodes amino acid sequence <SEQ ID 9880> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2759> which encodes the amino acid sequence <SEQ ID 2760>. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1958(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-1003-

Identities = 247/301 (82%), Positives = 274/301 (90%)

Query: 1 MRIQQLQYVIVIKIVETGSMNSAAKQLYITQPSLSNAVNRLSTEMGIQIRPNFKGITLTKD 60
 MRIQQL Y+IKIVE GSMNSAAKQL+ITQPSLSNAV++LE EMGI IF KNPKGITLTKD

5 Sbjet: 1 MRIQQLHVIKIVECGSMNSAAKQLFITQPSLENNAVKDLEMEMGITTFNPNFKGITLTKD 60

Query: 61 GMEFLSYARQILEQTALLEERYKGDNTSRELFSEVSSQHYAFVNVAFALFNGTDMTQYEL 120
 G+EFLSYARQI+EQT+LLE+RYK NT RELFSEVSSQHYAFVNVAF+L TDMT+YEL

10 Sbjet: 61 GVEFLSYARQILEQTALLEERYKNHNTGRELFSEVSSQHYAFVNVAFVSLKRTDMTRYEL 120

Query: 121 FLRETRTWELIIDVKNFRSEIGVLFLNSYNRDLVTLKLPDNDLSIATLTFTTTHPIFVSKS 180
 FLRETRTWELIIDVKNFRSEIGVL+N YNRDLVTLKLPDNR L A+ L F PHIFVSKS

15 Sbjet: 121 FLRETRTWELIIDVKNFRSEIGVLFINYNRDLVTLKLPDNDLSIATLTFTTTHPIFVSKS 180

Query: 181 NPLANKKLNKMDLEDYFYLSDQGLHNSFYFSEEMMSQIPHKSIVVSDRATLFLNMIG 240
 NPAA + L+M DL D+PYLSYDQG+HNSFYFSEEMMSQ+PH KSIVVSDRATLFLNMIG

18 Sbjet: 181 NPLATKLLSMDDLRDPTFYLSDQGLHNSFYFSEEMMSQMPHNKSIVVSDRATLFLNMIG 240

Query: 241 LDGYTAVTGILNSKLNQDEIVAIPLDVEDVIDIVIRHDKANLGRKMQQKFIYDLLEEVFN 301
 LDGYTVA+GLINS LNQD+TVAIFLDV D IDIV+I+H+KANLSRKG++FI+YLLEEV+F+

20 Sbjet: 241 LDGYTAVSGILNSKLNQDQVAIFLDVDEIDIVIKHEKANLSRKGSRFIEYLLEEVTFD 301

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 905

A DNA sequence (GBSx0960) was identified in *S. agalactiae* <SEQ ID 2761> which encodes the amino acid sequence <SEQ ID 2762>. This protein is predicted to be CpsX protein. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq

INTSGRAL	Likelihood	Transmembrane	22 - 38	(13 - 42)
INTSGRAL	Likelihood = -14.91	Transmembrane	52 - 68	(44 - 77)
INTSGRAL	Likelihood = -6.74	Transmembrane	76 - 92	(73 - 97)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.6965 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

40 The protein has homology with the following sequences in the databases:

>GP: AAC44935 GB: U6901 putative transcriptional regulator [Bacillus subtilis]
 Identities = 120/389 (30%), Positives = 196/389 (49%), Gaps = 17/389 (4%)

Query: 2 KIGKKIVMLFTAVLTIVLALGVLTSAVTFSTGELSKTFKDPSTSSNKSDAIK-QTRAF 60
 KI K+L+P A+ L V+ LG Y + E + S+ +++ + + F

45 Sbjet: 19 KILKRIMLLF-ALALLVVGVLGGYLYKTKINADESYDALGRKNKSNLRNEVDMKPKF 77

Query: 61 SILMGVDTGSSERASKNGNSDSMILVIVNPKTKTKTMTSLERDITLTLSGPKNNMG 120
 SIL MG+ + + + +G SDS+I+VT++PK K M S+ RDT L+G + G

50 Sbjet: 78 SILFMGIDYATKQ---KGRSDSLVVVLDPKNTKMKLSIPDTRVQLAG---IDTQ 130

Query: 121 VEAKINAAAYGASQAMTAVQDLNITIDYVQINMGQLIDLNAVAGGIVTNEDEPFI 180
 + K+NAAY+ GG + TV++ L I ID YV ++ G D+N VGGI V HGF

55 Sbjet: 131 SKTKINAAYSKGGKDEIVETVENFLQIPIDKYVTVDGDFGVDVNVGGIDVDVVPDFDE 190

Query: 181 SIAENEPYQATVAPGTHKINGEQALVYARMEDYDPEGYGRQKREVIQVLLKIAL 240
 +E + + G +NGE+AL YARMR D GD+GR RQ+ +++ + + +

18 Sbjet: 191 KSDVDESK-RITYPKGEMHNLGKALAYARMRKQKRGDFGRDRQQLNALDRKSSA 249

60 Query: 241 DSISSYKILSAVSSNMVNIETISSRTIPSLIGYRDALRTIKTYQLGDEATLSGGSYQ 300
 +I+ KI S N++TKI I+ + + I T + G D L +Y

-1004-

Sbjct: 250 SNIKIDKIAEKGSENVETNIRITEGLALQOIYSGFTSKKIDTSLTSGDLVGPNNITYY 309
 Query: 301 IVTSNHILLEIQNRIRPTELGLHKVNLQIKTNATVYENLYGSKTSQTVNNNYDSSGQAPSYSD 360
 LE ++R L H ++ +T T S + + + S+G +
 Sbjct: 310 FEPDATNLE---KVRKTLQEH-LDYTFDTSTGTSSTELGDTSSSSSGSGTGTGTTDGT 365
 Query: 361 SHSSYANYSSGVDTGQSASTDQDSTASSH 369
 + SSY+N SS T + ST +T SS+
 Sbjct: 366 NGSSYNDSS---TSSNNSTNSITDSSY 391

There is also homology to SEQ ID 2764.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 906

- 15 A DNA sequence (GBSx0961) was identified in *S.agalactiae* <SEQ ID 2765> which encodes the amino acid sequence <SEQ ID 2766>. This protein is predicted to be CpsIaB. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence
 20 INTEGRAL Likelihood = -0.75 Transmembrane 121 - 137 (121 - 137)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1298 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9877> which encodes amino acid sequence <SEQ ID 9878> was also identified.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 907

- 35 A DNA sequence (GBSx0962) was identified in *S.agalactiae* <SEQ ID 2767> which encodes the amino acid sequence <SEQ ID 2768>. This protein is predicted to be cpsb protein. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence
 40 INTEGRAL Likelihood = -9.02 Transmembrane 182 - 198 (179 - 204)
 INTEGRAL Likelihood = -5.57 Transmembrane 30 - 46 (24 - 46)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4609 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 45 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 10785> and protein <SEQ ID 10786> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 9

-1005-

McG: Discrim Score: -8.96
 GVH: Signal Score (-7.5): 0.11
 Possible site: 35
 >>> Seems to have no N-terminal signal sequence
 5 ALOM program count: 2 value: -9.02 threshold: 0.0
 INTEGRAL Likelihood = -9.02 Transmembrane 182 - 198 (179 - 204)
 INTEGRAL Likelihood = -5.57 Transmembrane 30 - 46 (24 - 48)
 PERIPHERAL Likelihood = 6.21 113
 modified ALOM score: 2.30
 10 *** Reasoning Step: 3
 ----- Final Results -----
 15 bacterial membrane --- Certainty=0.4609(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

20 Example 908

A DNA sequence (GBSx0963) was identified in *S.agalactiae* <SEQ ID 2769> which encodes the amino acid sequence <SEQ ID 2770>. This protein is predicted to be CpsIaD. Analysis of this protein sequence reveals the following:

Possible site: 61
 25 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.44 Transmembrane 149 - 165 (149 - 166)
 ----- Final Results -----
 30 bacterial membrane --- Certainty=0.1977(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 909

A DNA sequence (GBSx0964) was identified in *S.agalactiae* <SEQ ID 2771> which encodes the amino acid sequence <SEQ ID 2772>. Analysis of this protein sequence reveals the following:

Possible site: 25
 40 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -12.26 Transmembrane 276 - 292 (270 - 297)
 INTEGRAL Likelihood = -4.62 Transmembrane 10 - 26 (9 - 28)
 INTEGRAL Likelihood = -4.14 Transmembrane 41 - 57 (39 - 58)
 INTEGRAL Likelihood = -3.24 Transmembrane 100 - 116 (100 - 116)
 45 INTEGRAL Likelihood = -3.08 Transmembrane 445 - 461 (443 - 461)
 ----- Final Results -----
 50 bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8687> and protein <SEQ ID 8688> were also identified. Analysis of this protein sequence reveals the following:

-1006-

```

Lipop: Possible site: -1  Crend: 8
McG: Discrim Score:      5.69
GVH: Signal Score (-7.5): -5.63
    Possible site: 25
5  >>> Seems to have an uncleavable N-term signal seq
    ALOM program   count: 5 value: -12.26 threshold: 0.0
        INTEGRAL   Likelihood =-12.26  Transmembrane  276 - 292 ( 270 - 297)
        INTEGRAL   Likelihood = -4.62   Transmembrane  10 - 26 ( 9 - 28)
10      INTEGRAL   Likelihood = -4.14   Transmembrane  41 - 57 ( 39 - 58)
        INTEGRAL   Likelihood = -3.24   Transmembrane  100 - 116 ( 100 - 116)
        INTEGRAL   Likelihood = -3.08   Transmembrane  445 - 461 ( 443 - 461)
        PERIPHERAL Likelihood = 2.23    221
    modified ALOM score: 2.95
15  *** Reasoning Step: 3
    ----- Final Results -----
        bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
20      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 910

25 A DNA sequence (GBSx0965) was identified in *S.agalactiae* <SEQ ID 2773> which encodes the amino acid sequence <SEQ ID 2774>. This protein is predicted to be CpsF. Analysis of this protein sequence reveals the following:

```

    Possible site: 13
30  >>> Seems to have no N-terminal signal sequence
        INTEGRAL   Likelihood = -2.60   Transmembrane  79 - 95 ( 78 - 95)
    ----- Final Results -----
        bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
35      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 911

A DNA sequence (GBSx0966) was identified in *S.agalactiae* <SEQ ID 2775> which encodes the amino acid sequence <SEQ ID 2776>. This protein is predicted to be galactosyltransferase. Analysis of this protein sequence reveals the following:

```

    Possible site: 39
45  >>> Seems to have no N-terminal signal sequence
    ----- Final Results -----
        bacterial cytoplasm --- Certainty=0.4634(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
50      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.pyogenes*.

-1007-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 912

A DNA sequence (GBSx0967) was identified in *S.agalactiae* <SEQ ID 2777> which encodes the amino acid sequence <SEQ ID 2778>. Analysis of this protein sequence reveals the following:

```

Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
10  INTEGRAL    Likelihood =-12.47    Transmembrane    59 - 75 ( 54 - 82)
    INTEGRAL    Likelihood =-10.88    Transmembrane    309 - 325 ( 307 - 332)
    INTEGRAL    Likelihood =-8.07     Transmembrane    33 - 49 ( 28 - 53)
    INTEGRAL    Likelihood =-6.48     Transmembrane    195 - 211 ( 187 - 212)
    INTEGRAL    Likelihood =-6.16     Transmembrane    285 - 301 ( 283 - 306)
    INTEGRAL    Likelihood =-4.09     Transmembrane    222 - 238 ( 221 - 240)
15  INTEGRAL    Likelihood =-3.50     Transmembrane    78 - 94 ( 77 - 96)
    INTEGRAL    Likelihood =-2.71     Transmembrane    101 - 117 ( 99 - 117)
    INTEGRAL    Likelihood =-2.44     Transmembrane    8 - 24 ( 7 - 25)
    INTEGRAL    Likelihood =-1.59     Transmembrane    147 - 163 ( 147 - 164)
    INTEGRAL    Likelihood =-0.48     Transmembrane    168 - 184 ( 168 - 184)

20  ----- Final Results -----
        bacterial membrane --- Certainty=0.5969(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25  The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB43614 GB:AJ239004 polysaccharide polymerase [Streptococcus pneumoniae]
Identities = 74/309 (23%), Positives = 137/309 (43%), Gaps = 36/309 (11%)

30  Query: 53 FERRKLV---IIFLLFIATILNLFVHKVTFILTLIFFLAKDI--SLKKAFSIIIGSRI 107
    FE+RK II ++ I T+L + ++ +F+ I L++ II
    Sbjct: 61 FEKRKYTLQFIISIIITLLTYTICQMNYVYFTSWFMILIGTHYDLRRVKKIIFIVS- 119

    Query: 108 LGVLLNQIFVKLDLIEIKY----VNFYRDGQFILRSDLGFGHNFIHNFALITFLYIV 162
    L ++ IF+ L + I Y +N R+ + + GF HFN + ++I
35  Sbjct: 120 LSIHFISIFISILMYIIDYKREILNIRRN-ETVRAPTFGPIHFNKFTIVLSNLCNLPIN 178

    Query: 163 LNYKRLKPVMVFLFLNLYLQYTFSTRTGYIIVILFIVLYVTGNSLIKRVFMKLPAPY 222
    L RIK + L + Y +T +RT +I+ L+Y+ ++ ++ Y
40  Sbjct: 179 LKDRLEKYHVTFCLQIFLQFFYPTQRTALLSVIVIFALLYI--YMFVENLERIWIGYS 236

    Query: 223 QFFLLVFTFLSSTIFFNEN--FVQKLVLVLTGRHY-AHLQVLVDGLTFPGNSFKE----- 274
    F + F + + P+ SN F +D +LGR+ A+ + G T+G +
    Sbjct: 237 FFCISTFLGLVLAFOFYPGNKFSIPIDNHLGRILKLAAYARTTFGGYFWGQVQVKEIWD 296

45  Query: 275 -----TSVLFDSNSYMLLSMGYGVLTMFOMIIV-----YIYSKKIIIEQLLLFINSII 324
    TS FD+ YS L+S G++ + ++ Y+ +K +I+ LL+ M +
    Sbjct: 297 PIWGLTSFTDSDFYSLMSNAGIIMLLLSVLFLVKLQYLDNKSLL-----LLAWSMYAV 352

50  Query: 325 LFTEFSYPS 333
    T+ +PS
    Sbjct: 353 TETDLIFPS 361
  
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 913

A DNA sequence (GBSx0968) was identified in *S.agalactiae* <SEQ ID 2779> which encodes the amino acid sequence <SEQ ID 2780>. This protein is predicted to be cap8J. Analysis of this protein sequence reveals the following:

```

5       Possible site: 57
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3424 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GF:CAB43613 GB:AJ239004 cap8J [Streptococcus pneumoniae]
Identities = 94/237 (39%), Positives = 135/237 (56%), Gaps = 10/237 (4%)

Query: 1  MIPKVIHYCNFGNPLFDNLKKYIKTWREQCPDYHIEWNEHNYDVSKNVPMREAYTKKN 60
      MIPK IHY NFGG+ PD + K I +N++ EDYEL-ENNE N+D+S + F + AY +
Sbjct: 1  MIPKKIHYNFGGSEKEDVVLKICINSWCKMPDYHIEWNEHNFDSQFPAKSAYESRK 60

Query: 61  FAYVSDYARLDIIYTYGGFYLDTVLKLKSL-DPLRIHECFIAREISCDVNTGLIIGAVK 119
      +A+ SDYAR I+ YGG Y DTDVLLK+ D + H F R +VN GL+ +
Sbjct: 61  WAFASDYARFKILSKYGGIYFDTVELLKTISDDILAHSSFTGPEYIGEVNPLGVYACMP 120

Query: 120  GHHFLKSNMSTYDKS--DLTSLNKTCEVITNLLINRGLKQNWIIQKIDDDITYPRNFTN 177
      K + Y+++ D+ L T +T+ L+ + N Q ID + IYP +YF
Sbjct: 121  DDKIAKYMVQYEQASFDINHL-VTVNTIITDYLLQNFNQFNQFIIDGLAIYEDDYFVC 179

Query: 176  PKNLLTKRVDCTSVTYSIHHYEGSKSSPISDLAKIRVLIIDFLRGYGYTRMLL 234
      + +V LT T SIHHY +WK+ +LK +V++I+ + G YR LL
Sbjct: 180  GYDQVKEVR-LTERTISIHYSATWKR-----TLKRIVQMIVKTIIGAEYRKLL 230

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 914

A DNA sequence (GBSx0969) was identified in *S.agalactiae* <SEQ ID 2781> which encodes the amino acid sequence <SEQ ID 2782>. Analysis of this protein sequence reveals the following:

```

40      Possible site: 44
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
45      bacterial cytoplasm --- Certainty=0.3897 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GF:CAAS7700 GB:Z47767 WbcL [Yersinia enterocolitica]
Identities = 60/207 (28%), Positives = 101/207 (47%), Gaps = 22/207 (10%)

Query: 4  IFITPTFNRGYRLSYLYDLSLCNITKNFIMLIVDDGSEDSTKEIVSNYIKENKVSIVLYLK 63
      +FTPTFNR + L Y S+ Q + WLIVDDGS D+T E+V ++ ENK+I Y+Y+
Sbjct: 6  VFTPTFNRAHVLKRCYLSILEQDRDDIEMLIVDDGSTNTARVDSFKNENKLNKYIYQ 65

Query: 64  RNGGKHSAYNLAMRYMQPSDYHVCVDSIDWLEDAV-----EIIKDLDELSTLSNRVYG 117
      N GK +A+N A+ +Y + +DSDD + ++ +F D E + +

```

-1009-

Sbjct: 66 DNSGKQAARKAVENAS-GEYFIGLSDDAFLAGSINKLLSMNAVDFDKRIIGIR----A 120
 Query: 118 LVPYPRYSLNQGRNINLPKILEVNIPLACKYHLEKIEICVINNAVLDVEFPCTGEGNPL 177
 + +L N +L+ + + + D + + + R L + +P G N P +
 Sbjct: 121 ISVSSETLKPNNTYLSNEDKKSSWFD-BFSSGIRGERIDFFKTELLRKLYLPVASGINFI 179
 Query: 178 SEEMWYILSKKGYFCPQNRKIYCFDY 204
 E Y ++K+ YCF Y
 Sbjct: 180 PEINWFYSTVAKS-----YCFY 196

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 915

A DNA sequence (GBSx0970) was identified in *S.galactiae* <SEQ ID 2783> which encodes the amino acid sequence <SEQ ID 2784>. This protein is predicted to be eps7. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -2.18 Transmembrane 190 - 206 (189 - 206)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB59293 GB:AJ131984 putative galactosyl transferase
 [Streptococcus pneumoniae]
 Identities = 101/312 (32%), Positives = 172/312 (54%), Gaps = 4/312 (1%)
 Query: 3 LISIIVPVYNGEYIGRCLDSILEQTYNLEIIIDDGSSDRDTGDCIKYFLDDRRIKYP 62
 +IS+IVFVYN Y+ LDS+LEQTY++ E+I++DGS+D +G+IC++Y I F
 Sbjct: 1 MISIVFVYNVADYLRFAISLLRQTYKDFEVLVNDGSTNDSGEICDEYGLKLYDNIHVF 60
 Query: 63 YQENRQGSVARNNGVLRCTGDNIAPLSDVLPYSIEVMYNIQKATNADIVLT--SIGN 120
 ++N G S ARN G+ + G++I FLSDSD + FY++E++ IQK + DIV T I
 Sbjct: 61 HKKNGSLDARNFGLKSRGEPTVFLSDDYFRPVALELLITQKQYDVIDVSTWGGITY 120
 Query: 121 PNTYNTSINSQYLKRIKLYLTLEVALEMYGKTYGVSPKLKLYPRSNLLSNPYFEGKI 180
 ++ Y+ + ++ +K+ T + L +YY VS KLY R +L +P+GKI+
 Sbjct: 121 SHDIYSKKLMAEDYLYIKLITKCEFLAAVYNDENFTSWAGKLYKR-DLFTKTFPGKIY 179
 Query: 181 EDMDITPKLISCASKIAVCDIVTAVVYFSDNSTRTKPNRMGLYFPAIGNVIFININP 240
 ED+ + + +A D+ Y S + F++R PF+AI +N I +
 Sbjct: 180 EDLYVVAERLLNIKTVAHFDLPIYHYQRQGSIVNSTFSDRQYDFDPAIDINEALIKFY 239
 Query: 241 PHNTSLISAVIYNEVPGGIDIGSGLDKLYDVTVDYRKYRKYFTTILPNNRISVKEV 300
 + L++A+ V G I + + + + Y+ ++ N +I +K KV
 Sbjct: 240 GSDKELLAALNAKRVIGSF-ILNSAFYNSKNDITKIRIKFYWEVINKKIKPMKRKV 298
 Query: 301 KYILFISSRYF 312
 + +LF+ S Y+
 Sbjct: 299 CQVLFLLSPNTY 310

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1010-

Example 916

A DNA sequence (GBSx0971) was identified in *S.agalactiae* <SEQ ID 2785> which encodes the amino acid sequence <SEQ ID 2786>. This protein is predicted to be galactosyltransferase. Analysis of this protein sequence reveals the following:

```

5      Possible site: 26
      >>> Seems to have an uncleavable N-term signal seq

      ----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
10     bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2787> which encodes the amino acid sequence <SEQ ID 2788>. Analysis of this protein sequence reveals the following:

```

15     Possible site: 28
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2065 (Affirmative) < succ>
20     bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

25     Identities = 37/111 (33%), Positives = 61/111 (54%), Gaps = 3/111 (2%)

Query: 1   MDKVSIIIPVYNVQSFNLECTIESVLAQ-TYSNLEILLVMDGSTDNGDIC-DYYSIDGR 58
          M KVSII  YN  +++ ++S L+Q T  +EII++D STD+S +I  Y  +  G+
Sbjct: 1   MYKVSIICTNYNKAPWISDALDSFLSQVTD+FEVEIIVDDASTDSREILKSYQKSSGK 60

30     Query: 59  I-FVPHKNNGLSDARNYGISRATGDYIYLLSDSDYLYKEDAIERMVFSE 108
          I  +F++ N G++      A G YI  D DDY      +++ V+  E
          Sbjct: 61  IKLLFNETNIGITKTWIKACLYAGSKYLARCDGDDYWTDSFKLQKQVVL 111

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 917

A DNA sequence (GBSx0972) was identified in *S.agalactiae* <SEQ ID 2789> which encodes the amino acid sequence <SEQ ID 2790>. This protein is predicted to be CpsK. Analysis of this protein sequence reveals the following:

```

40     Possible site: 52
      >>> Seems to have an uncleavable N-term signal seq

      ----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
45     bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 918

A DNA sequence (GBSx0973) was identified in *S.agalactiae* <SEQ ID 2791> which encodes the amino acid sequence <SEQ ID 2792>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 31
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1956(Affirmative) < succ>
10     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 919

A DNA sequence (GBSx0974) was identified in *S.agalactiae* <SEQ ID 2793> which encodes the amino acid sequence <SEQ ID 2794>. This protein is predicted to be capsular polysaccharide. Analysis of this protein sequence reveals the following:

```

20     Possible site: 36
      >>> Seems to have an uncleavable N-term signal seq

      INTEGRAL Likelihood = -8.81 Transmembrane 89 - 105 ( 80 - 112)
      INTEGRAL Likelihood = -7.01 Transmembrane 439 - 455 ( 428 - 460)
25     INTEGRAL Likelihood = -5.74 Transmembrane 322 - 338 ( 317 - 342)
      INTEGRAL Likelihood = -4.88 Transmembrane 175 - 191 ( 174 - 195)
      INTEGRAL Likelihood = -3.45 Transmembrane 146 - 162 ( 145 - 166)
      INTEGRAL Likelihood = -3.08 Transmembrane 381 - 397 ( 375 - 398)
      INTEGRAL Likelihood = -2.50 Transmembrane 413 - 429 ( 412 - 430)
30     INTEGRAL Likelihood = -1.91 Transmembrane 206 - 222 ( 205 - 222)
      INTEGRAL Likelihood = -1.59 Transmembrane 354 - 370 ( 354 - 372)
      INTEGRAL Likelihood = -1.54 Transmembrane 43 - 59 ( 43 - 61)
      INTEGRAL Likelihood = -0.22 Transmembrane 252 - 268 ( 252 - 268)

      ----- Final Results -----
35     bacterial membrane --- Certainty=0.4524(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 920

A DNA sequence (GBSx0975) was identified in *S.agalactiae* <SEQ ID 2795> which encodes the amino acid sequence <SEQ ID 2796>. This protein is predicted to be NeuB. Analysis of this protein sequence reveals the following:

```

45     Possible site: 30
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
50     bacterial cytoplasm --- Certainty=0.2992(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

```

-1012-

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 921

A DNA sequence (GBSx0976) was identified in *S.agalactiae* <SEQ ID 2797> which encodes the amino acid sequence <SEQ ID 2798>. This protein is predicted to be NeuC. Analysis of this protein sequence reveals the following:

```
10 Possible site: 41
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.3150(Affirmative) < succ>
15 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 922

A DNA sequence (GBSx0977) was identified in *S.agalactiae* <SEQ ID 2799> which encodes the amino acid sequence <SEQ ID 2800>. This protein is predicted to be NeuD. Analysis of this protein sequence reveals the following:

```
25 Possible site: 16
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

There is homology to SEQ ID 542.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 923

A DNA sequence (GBSx0979) was identified in *S.agalactiae* <SEQ ID 2801> which encodes the amino acid sequence <SEQ ID 2802>. Analysis of this protein sequence reveals the following:

```
40 Possible site: 33
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.2576(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

-1013-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 924

A DNA sequence (GBSx0980) was identified in *S.agalactiae* <SEQ ID 2803> which encodes the amino acid sequence <SEQ ID 2804>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1621(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9875> which encodes amino acid sequence <SEQ ID 9876> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2805> which encodes the amino acid sequence <SEQ ID 2806>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1066(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

Identities = 83/139 (59%), Positives = 111/139 (79%)

```
Query: 6 TETHDHOALIQKLLVSIHYLTLPFDEIILVEKTPSLGKHFSIAIVQNELGSLISKIAL 65
TE + HQ LIQKLLVSIHYLTLPFDE+ LVE+TPS+LG F +VQ+ELG+I++ I+ L
Sbjct: 4 TQNSHQILIQKLLVSIHYLTLPFDEKLIVERTPSILQGEFFPAHLVQSELGDIVAIDTL 63

Query: 66 SKQKKLIRSIWYDESSFKVMKALAIVEEMIKGLDNLLEFCQSQTVPQAILGDERAHVF 125
Q++LI S +WY+ES+FK+MK L IV+ WIKG+D+L++ QGS+ VFQ I+GD+R VF
Sbjct: 64 DMQQRLLIRSTFWYEEAFKLMNKLTLDFVDNWIKGVDLIDLCQSKGVFIITIGDKIRIVF 123

Query: 126 GILIDVYTSNIINFTSLKE 144
G+L DV++SL + SLKE
Sbjct: 124 GVLSDVFSSLAIVSALSKE 142
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 925

A DNA sequence (GBSx0981) was identified in *S.agalactiae* <SEQ ID 2807> which encodes the amino acid sequence <SEQ ID 2808>. This protein is predicted to be uracil-DNA glycosylase (ung). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3427(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2809> which encodes the amino acid sequence <SEQ ID 2810>. Analysis of this protein sequence reveals the following:

```

Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4200 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 160/216 (74%), Positives = 185/216 (85%)

Query: 1  MKHSSMDLIKRELPNHYYNKINTFMADVYSGIYVPRDKVFNAQITPLENKVVIIG 60
      M HS NH+ IK LP HYY +IN F+D Y SG+VYPPR+ VF A+Q+TPLE KV+I+G
Sbjct: 1  MAHSINHEKIKFLEPNHYGRINHLDEAYASGLVYPPRENVFKALQVTPLEETKVLIIG 60

Query: 61  QDPYHGPGQAQGLSFSVPNLFAPPSLQNLKELAEIDGSRSHDLTSNAQQGVLLNLNAC 120
      QDPYHGP+QAQGLSFSVP+ + APPSL NIIKELA+DIG R HNDL++WA QGVLLNLNAC
Sbjct: 61  QDPYHGPGQAQGLSFSVPBISAPPSLINILKELADIDGPRDHDLSWASQGVLLNLNAC 120

Query: 121  LTVPEHQANGHAGLINEPFTDAVIKVVNQKSTPVVFIWGGYARKKKSLIDNPPIHIIIES 180
      LTVP QANGHAGLINEPFTDAVIKVN+K+PVVFIWLG YARKKK+ I NP HHIIIES
Sbjct: 121  LTVPAGQANGHAGLINEPFTDAVIKVNKEKSPVVFIWQAYARKKKAFITNPPIHIIIES 180

Query: 181  PHPSPLSAYRGFFGSRPFPSRTNHFLEEGINEIDWL 216
      PHPSPLS+YRGFFGS+PFPSRTN LE+EG+ +DWL
Sbjct: 181  PHPSPSSYRGFFGSKPFSRTNAILEKEGMTGVDWL 216

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 926

A DNA sequence (GBSx0982) was identified in *S.agalactiae* <SEQ ID 2811> which encodes the amino acid sequence <SEQ ID 2812>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -11.15 Transmembrane 147 - 163 ( 109 - 166)
INTEGRAL Likelihood = -8.92 Transmembrane 124 - 140 ( 109 - 146)
INTEGRAL Likelihood = -6.16 Transmembrane 167 - 183 ( 166 - 186)
INTEGRAL Likelihood = -4.67 Transmembrane 3 - 19 ( 1 - 23)
INTEGRAL Likelihood = -3.98 Transmembrane 72 - 88 ( 64 - 92)
INTEGRAL Likelihood = -1.06 Transmembrane 106 - 122 ( 105 - 122)
INTEGRAL Likelihood = -0.90 Transmembrane 54 - 70 ( 54 - 70)

----- Final Results -----
      bacterial membrane --- Certainty=0.5458 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9873> which encodes amino acid sequence <SEQ ID 9874> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA91549 GB:Z67739 unidentified [Streptococcus pneumoniae]
Identities = 134/212 (63%), Positives = 168/212 (79%)

Query: 1  MNIIIMIIAYLLGSIQTGLWIGKPYQVNLRCGSGNTGTNTFRILGVKGIIVLTID 60

```


-1015-

M I+++IAYLLGSI +GLWIG+ P+Q+NLR+HSGSGNTGTTNTRILG KAG+ T ID
 Sbjct: 1 MITIVLLILAYLLGSI P+GLWIGQVFPQINLRHSGSGNTGTTNTRILGKAGMATFVID 60
 Query: 61 ILKGTLATLPIIILGITTVPFFIGFPAIIGHTFFPIPAQKGGKAVATSAGVLLGFAPSF 120
 KGTAL+PLII + VSP G A+IGHTFFPIFA PKGGKAVATSAGV+ GFAP F
 Sbjct: 61 FFKGTLATLPIIIFILQGVSPILFGLLAVIGHTFFPIFAGPKGGKAVATSAGVIFGFAPIF 120
 Query: 121 FLYLVLPILTLTYLFMSISLSSITVAVVGILSVLFPVLGFIITDYDMITPTTVVILMAIT 180
 LYL +IF LYL SMISLS+T ++ ++ VL+PFL GFLA++YD+P +++ +A
 Sbjct: 121 CLYLAIFFGALYLGSMISLSSVTASIAVIGVLLFPLFGFILSNYDPLFIATILALASL 180
 Query: 181 IIRHQNRIKRIKRIKQENLVPGLNLKQPK 212
 IIRH+DNE RI+ + EMLVP+GLNL+ Q K
 Sbjct: 181 IIRHKNRIKRIKRIKQENLVPGLNLTHQPK 212

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2813> which encodes the amino acid sequence <SEQ ID 2814>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -10.83 Transmembrane 194 - 210 (191 - 216)
 INTEGRAL Likelihood = -9.77 Transmembrane 146 - 162 (132 - 191)
 INTEGRAL Likelihood = -7.70 Transmembrane 165 - 181 (163 - 191)
 INTEGRAL Likelihood = -5.89 Transmembrane 23 - 39 (19 - 47)
 INTEGRAL Likelihood = -4.83 Transmembrane 95 - 111 (91 - 118)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GB:CAA91549 GB:Z67739 unidentified (Streptococcus pneumoniae)
 Identities = 138/213 (64%), Positives = 166/213 (77%)
 Query: 28 MKLLPITTIAYLLGSIPTGLWIGQFYFHNLRHSGSGNTGTTNTRILGKAGTATLAID 87
 M ++ + +AYLLGSI +GLWIGQ F+ INLRHSGSGNTGTTNTRILG KAG AT ID
 Sbjct: 1 MITIVLLILAYLLGSI P+GLWIGQVFPQINLRHSGSGNTGTTNTRILGKAGMATFVID 60
 Query: 68 MPKGTLSILLPIIPGMTSISLSIAIGFFAVLGHFTFFIPANPKGGKAVATSAGVLLGFAPLY 147
 FKGTL+ LLIPII + +S + G A+VGHFTFFIFA PKGGKAVATSAGV+ GFAP++
 Sbjct: 61 FFKGTLATLPIIIFILQGVSPILFGLLAVIGHTFFPIFAGPKGGKAVATSAGVIFGFAPIF 120
 Query: 148 LFFLASIFVLVLYLFMSISLASVVSIAVGVLSVLTFAIHFLPLPNYDYFLTPTVILLAFI 207
 +LA IF LYL SMISL+SV +I V+ VL PP F+L NYD+ I++ LA +
 Sbjct: 121 CLYLAIFFGALYLGSMISLSSVTASIAVIGVLLFPLFGFILSNYDPLFIATILALASL 180
 Query: 208 IIRHKNDI+RIK+ TENL+PGLNL+ Q PKK
 IIRHKNDI+RIK+ TENL+PGLNL+ Q PKK
 Sbjct: 181 IIRHKNRIKRIKRIKQENLVPGLNLTHQPK 213

An alignment of the GAS and GBS proteins is shown below.

Identities = 143/212 (67%), Positives = 174/212 (81%)
 Query: 1 MNIIIMIIAYLLGSIQTLGWIGKYFYQVNLRHSGSGNTGTTNTRILGKAGVILTID 60
 M +++ I IAYLLGSI TGLWIG+YFY +NLR+HSGSGNTGTTNTRILGKAG TL ID
 Sbjct: 28 MKLLPITTIAYLLGSIPTGLWIGQFYFHNLRHSGSGNTGTTNTRILGKAGTATLAID 87
 Query: 61 ILKGTLATLPIIILGITTVPFFIGFPAIIGHTFFPIPAQKGGKAVATSAGVLLGFAPSF 120
 +KGTL+ L+PII G+T++S IGFFA++GHTFFPIFA PKGGKAVATSAGVLLGFAP+
 Sbjct: 88 MPKGTLSILLPIIPGMTSISLSIAIGFFAVLGHFTFFIPANPKGGKAVATSAGVLLGFAPLY 147
 Query: 121 FLYLVLPILTLTYLFMSISLSSITVAVVGILSVLFPVLGFIITDYDMITPTTVVILMAIT 180
 +L IF+L LYLFSMISL+S+ A+VH+SVL FP + F+L +YD+ T +VIL+A
 Sbjct: 148 LFFLASIFVLVLYLFMSISLASVVSIAVGVLSVLTFAIHFLPNYDYFLTPTVILLAFI 207

-1016-

Query: 181 IIRHQDNIRKIRKQENLVPPGIANLSKQKNK 212
 IIRH+DNI RI+ ENL+P+GLNLSKQ K
 Sbjct: 208 IIRHKDNLSRIKHTENLIPWGLNLSKQVPK 239

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 927

A DNA sequence (GBSx0983) was identified in *S. agalactiae* <SEQ ID 2815> which encodes the amino acid sequence <SEQ ID 2816>. Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane ---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside ---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm ---	Certainty=0.0000 (Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 928

A DNA sequence (GBSx0984) was identified in *S. agalactiae* <SEQ ID 2817> which encodes the amino acid sequence <SEQ ID 2818>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.1585 (Affirmative)	< succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside ---	Certainty=0.0000 (Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 9871> which encodes amino acid sequence <SEQ ID 9872> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GF:CAA91550 GB:267739 DNA topoisomerase IV [Streptococcus pneumoniae] (ver 2)
 Identities = 574/649 (88%), Positives = 617/649 (94%), Gaps = 2/649 (0%)

Query: 5 LAKQDITVTVNGDGAIQVLEGLDAVRKRPFGMYIGSTDGRGLHILVWEIVDNAVDEALSGF 64
 ++K+I + NY DDAIQVLEGLDAVRKRPFGMYIGSTD GRGLHILVWEIVDNAVDEALSGF
 Sbjct: 1 MSKKEININYNDDAIQVLEGLDAVRKRPFGMYIGSTDGRGLHILVWEIVDNAVDEALSGF 60

Query: 65 GNRIDVVIINKDGSITVTHNRCGMPTCMHAWKKPTVEVIFTVLHAGCKFGQGGYKTSGGHL 124
 G+RIDV INKDS+TV DGRGKPTGMHAWG PTEVEVIFT+LHAGCKFGQGGYKTSGGHL
 Sbjct: 61 GDRIDVVIINKDGSITVQDHRCGMPTCMHAWGIPTEVEVIFTLHAGCKFGQGGYKTSGGHL 120

Query: 125 GVGSVVYVALSSNLVEVILRDGAIVYRQFENGKGPVTTLLKKIGTAPKSKGTSVSPMPDQ 184
 GVGSVVYVALSSNLVEVIRDGA-Y+YRQFENGKGPVTTLLKKIGTAPKSK+GT V+YMPD
 Sbjct: 121 GVGSVVYVALSSNLVEVITRDGAIVYRQFENGKGPVTTLLKKIGTAPKSKGTGVTFMPDA 180

Query: 185 SVFSTIDFKNFTIAERLKSAPLLKNVTILTDNRSEAEHLEFHYENGQDFVEYIARD 244
 ++FST DFK+NTI+ERL ESAPLLKNVTIL+LTD R++EA +EFHYENGQDFV YIARD

-1017-

Sbjct: 181 TIPSTTDKFKNTISERLINESAPLLKGVTLSTDKTDEA--IEPHYENGQVQDFVSYLNE 238

Query: 245 KETLTPIMEFEGEQRSPHIEVALQYNDGFSNILLSFVNNVRTKDGTHETGLKSAITKM 304
KE LTP++FEGE+ F +EVALQYNDGFSNILLSFVNNVRTKDGTHETGLKSAITK M

5 Sbjct: 239 KEILTPVLYFEGEDNGFQVEVALQYNDGFSNILLSFVNNVRTKDGTHETGLKSAITKVM 298

Query: 305 NDYARKTGLLKEKDKNLGSDYREGLSAILSLVPEHILQFEGQTKDKLGSPLARIPVDG 364
NDYARKTGLLKEKDKNLGSDYREGI+A+LSLIVPEHILQFEGQTKDKLGSPLAR+VDC

10 Sbjct: 299 NDYARKTGLLKEKDKNLGSDYREGLAVALSLVPEHILQFEGQTKDKLGSPLARPVVDG 358

Query: 365 IVSEKLTPTFLMENGDLASNLIRKAIKARDAREARKARDESNGKKSKDKGLSGKITP 424
IV++KLT+FLMENG+LASNLIRKAIKARDAREARKARDESNGKK+KDKGLSGKITP

15 Sbjct: 359 IVADKLTPTFLMENGDLASNLIRKAIKARDAREARKARDESNGKKSKDKGLSGKITP 418

Query: 425 AQSKNKNELYLVEGDSAGSQAQGRDKFQAILPLRGKVINTAKAMADILKNEINT 484
AQSKN KNELYLVEGDSAGSQAQGRDKFQAILPLRGKV+NTAKAMADI+KNEINT

20 Sbjct: 419 AQSKNPAKNELYLVEGDSAGSQAQGRDKFQAILPLRGKVINTAKAMADILKNEINT 478

Query: 485 MIHTTGAGVGDFNLDINFDKIIMTDAITDGAHITLLTFFPYRMRPLVEGHVYIA 544
MI+TIGAGVG DF++D NYDKIIMTDAITDGAHITLLTFFPYRMRPLV GHVIA

25 Sbjct: 479 MIYTTGAGVGADFSIEDANYDKIIMTDAITDGAHITLLTFFPYRMRPLVEGHVYIA 538

Query: 545 LPPLYRSGKSGKKEIVAYMTDIELEELRQFGKGLQRYKGLSEMANADQLWETWNP 604
LPPLYRSGKSGKKE V YAMTD ELSELR++PFGH+ LQRYKGLSEMANADQLWETWNP

30 Sbjct: 539 LPPLYRSGKSGKKEIVAYMTDIELEELRQFGKGLQRYKGLSEMANADQLWETWNP 598

Query: 605 ETRTLIRVTIEDLARAERKRVNLMGDKVPRRQWIEKNVKTLEENIVF 653
ETRTLIRVTIEDLARAERKRVNLMGDKV PR+WIENNVKTLEE TVF

35 Sbjct: 599 ETRTLIRVTIEDLARAERKRVNLMGDKVPRRQWIEKNVKTLEENIVF 647

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2819> which encodes the amino acid sequence <SEQ ID 2820>. Analysis of this protein sequence reveals the following:

Possible site: 53
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.1518 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

40

An alignment of the GAS and GBS proteins is shown below.

Identities = 560/649 (86%), Positives = 615/649 (94%)

45 Query: 5 LAKQDITVINYGDAAIQVLEGLDAVRKRPOMYIGSTDGTGLHLHLWEIVDVADEALSGF 64
L K+IT+ NY DDAIQVLEGLDAVRKRPOMYIGSTD GTGLHL+WEIVDVADEALSGF
Sbjct: 2 LTKKEITVINYNDAAIQVLEGLDAVRKRPOMYIGSTDGTGLHLHLWEIVDVADEALSGF 61

Query: 65 GNRIDVILNKDGSITVTDKRMGMPQIMHMGKPTVIEVITPLHAGQKPGQGYKTSGLH 124
G + I V+INKDGS+V D GNMGPT HWAG PTV+VIFT+LHAGQKPGQGYKTSGLH
50 Sbjct: 62 GDDIRVINKDGSVSVADSGRMPTQGHAMGIPTVQVIFTPLHAGQKPGQGYKTSGLH 121

Query: 125 GVSSVUNALSSWLEVEIRIDGAIYRQFENGCKPVITLKKIGTAPKSGTSPVSMFDQ 184
GVSSVUNALS+WELEVEI RDG+YRQFENGCKPVITLKK+GTAPKSGT V+PMFD
55 Sbjct: 122 GVSSVUNALSAWLEVEITRDGSVYRQFENGCKPVITLKKVGTAPKSGTSPVSMFD 181

Query: 185 SVFSTIDPKFNTIARLAKESAPLLKNVTLTLDNRSEAEHLPEPHYENGQVDFVSYLNE 244
+FSTIDPKFNTI+ERLAKESAPLLKNV ++LTD R++ EPHYENGQVDFVSYLNE
Sbjct: 182 KIPSTIDPKFNTISERLAKESAPLLKNVMSLTDLRGDPIIEPHYENGQVDFVSYLNE 241

60 Query: 245 KETLTPIMEFEGEQRSPHIEVALQYNDGFSNILLSFVNNVRTKDGTHETGLKSAITKM 304
KETLTP++ EG+Q+F +EVALQYNDGFSNILLSFVNNVRTKDG+HETGLKSAITK+M
Sbjct: 242 KETLTPVIYMEGQDQDFQVEVALQYNDGFSNILLSFVNNVRTKDGSHETGLKSAITK+M 301

65 Query: 305 NDYARKTGLLKEKDKNLGSDYREGLSAILSLVPEHILQFEGQTKDKLGSPLARIPVDG 364
NDYARKT LLKKEKDKNLGSDYREGLSA+LSLIVPE+HLQFEGQTKDKLGSPLARIPV+

-1018-

Sbjct: 302 NDYARKTNLLEKDKNLGGSDYREGLSAVLSILVPEQHLOPBGQTKDKGLSPARPIVES 361
 Query: 365 IVSEKLTLYFLMENGDLACNLIRKAIKARDAREARKARDSENRNGKSKEDKGLLSGKLTTP 424
 IVSEKLT+FL+ENG++AS+L+RKAIKARDAREARKARD+SRNGSK+KEDKGLLSGKLTTP
 Sbjct: 362 IVSEKLTFFLLENGEVASHLVKRAIKARDAREARKARDSENRNGKKNKDKGLLSGKLTTP 421
 Query: 425 AQSKNAKKNELLYLVBGDSAGGSAKQGRDRKFOAILPLRGKVLNTAKAMADIKNKEEINT 484
 AQSKNAKKNELLYLVBGDSAGGSAKQGRDRKFOAILPLRGKVLNT KAKMADI+KNEEINT
 Sbjct: 422 AQSHQAKKNELLYLVBGDSAGGSAKQGRDRKFOAILPLRGKVLNTEKAKMADILKNEEINT 481
 Query: 485 MIHTIGAGVGPDPNLDDINYSKIIIMTDADTGAHIQ/LLLTFFPYRMRPLVEBGHVYIA 544
 M++TIIGAGV GPDL+DINYSKIIIMTDADTGAHIQ/LLLTFFPYRMRPLVE BGHVYIA
 Sbjct: 482 MVTTIGAGVGADPNLEDINYSKIIIMTDADTGAHIQ/LLLTFFPYRMRPLVEAGHVYIA 541
 Query: 545 LPPLVYMSKGGKKEIVEYAWTDIELESLRQKFGKGLLQRYKGLGBMNADQLMETTMNP 604
 LPPLVYMSKGGK E + YAWTD ELS+LR++FGKG++LQRYKGLGBMNA+QLMETTM+P
 Sbjct: 542 LPPLVYMSKGGKTSKIAVWTDCELEDLREFGKQAILQRYKGLGBMNANQLMETTMNP 601
 Query: 605 ETRTLIRVTIEDLARAERRVYVLMGDKVPRRQWIEDNVKFTLEBNTVF 653
 ETRTLIRVTI+DLARAERRV+VLMGDK PRRWIEDNVKFTLEBNTVF
 Sbjct: 602 ETRTLIRVTIIDLARAERRVYVLMGDKVAPRRQWIEDNVKFTLEBNTVF 650

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 929

A DNA sequence (GBSx0985) was identified in *S. agalactiae* <SEQ ID 2821> which encodes the amino acid sequence <SEQ ID 2822>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence
 30 INTEGRAL Likelihood = -0.80 Transmembrane 378 - 394 (378 - 394)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1319 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 35 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:RAAD34369 GB:AP129764 ParC [Streptococcus mitis]
 Identities = 640/820 (78%), Positives = 722/820 (88%), Gaps = 5/820 (0%)
 40 Query: 1 MSNIQMSLEDIMGERFGYSKYIIQERALPDIRDGLKPVQRRLYSMMKDGNTFEGFR 60
 MSNIQMSLEDIMGERFGYSKYIIQ+RALPDIRDGLKPVQRRLYSMMKDGNTF+K +R
 Sbjct: 1 MSNIQMSLEDIMGERFGYSKYIIQERALPDIRDGLKPVQRRLYSMMKDGNTFEGFSYR 60
 45 Query: 61 KSAKSVGNMGNFPHGDSIIYDMVMWSQDWKNRETLIEMHGNNGSMDDGPPAAMRYTE 120
 KSAKSVGN+MGNFPHGDSIIYDMVMWSQDWKNRE L+EMHGNNGSMDDGPPAAMRYTE
 Sbjct: 61 KSAKSVGNMGNFPHGDSIIYDMVMWSQDWKNRETLIEMHGNNGSMDDGPPAAMRYTE 120
 50 Query: 121 ARLSEIAGYLLQIDIKNTVPFANNPDOTEKRPFTVLPAPFPHLLVNGATGISAGYATDIPP 180
 ARLSEIAGYLLQIDIK TVPF+NNPDOTEKRPFTVLPAPFPHLLVNG+TGISAGYATDIPP
 Sbjct: 121 ARLSEIAGYLLQIDIKKTVPFNNPDOTEKRPFTVLPAPFPHLLVNGSTGISAGYATDIPP 180
 55 Query: 181 HNLAEVIDAVVYMDHFKAKLQKMEFLGDPDFTGAI IQGDEIRKAYETGKGRVAVRS 240
 HNLAEVIDA VYMDHP AK+DKLMEFLGDPDFTG IIQG+DEI+KAYETGKGRV VRS
 Sbjct: 181 HNLAEVIDAVVYMDHPTAKVDKLMEFLGDPDFTGGI IQGDEIRKAYETGKGRVVRS 240
 Query: 241 RTAIEFLKGGKQIIVTIPYEVNKSVLKRIIDVVRNNKVPGEIAVRDESDRDLGLRIAT 300
 +T IE LKGGK+QI++TIPYEVNK+ LVK+IDVVRN+KV GIEA VRDESDRDLGLRIAT
 Sbjct: 241 KTEIEFLKGGKQIVTIPYEVNKANLVKIIDVVRNKSVAIEA VRDESDRDLGLRIAT 300
 60 Query: 301 ELKKEADKTIIVNLYLKYITDLQVNNFNMVAIDDYTPKQVGLSRILTSYIAHRRRIIAR 360
 ELKK+A+ +VNLVLYKYITDLQ+VNNFNMVAID++TP+QVG+ ILSYIAHRRB+I+AR

-1019-

- 5
10
15
20
25
30
35
40
45
- Sbjct: 301 ELKKDANTELWLNLYLKYCTDLQINNFNMVAIDNFTPRQGVIVPILSSYIAHREVEILAR 360
 Query: 361 SFYDKGKAEKRLHIVBGLIRVLSILDEVIALIRASGNKADAKENLKVSYPESEAQAERIV 420
 S+FDKSKAEKRLHIVBGLIRVLSILDEVIALIRASGNKADAKENLKVSYP+FE QAEARIV 420
 Sbjct: 361 SRFDKSKAEKRLHIVBGLIRVLSILDEVIALIRASGNKADAKENLKVSYPDFTPEQAERIV 420
 Query: 421 TLQLYRLNTIDIVTLAESEELAQQTIMLKAILISDERTMYNMKRELREVKKKFANTRRS 480
 TLQLYRLNTID+V L+SEE ELR++I ML AII DERTMYN+MK+ELREVKKFA R S 480
 Sbjct: 421 TLQLYRLNTIDVVVLQEEAELEKILAMLAATIGDERTMYNLMKKELREVKKKFATPRLS 480
 Query: 481 ELQELAEITIDTASLIASEDTYVSVTGQYVKRTSPGPNASTVDELGRNDELIPVS 540
 L++ A+ IRIDTASLI SEDTYVSVT+ GY+KRTSPGPF AST++E+GKR+DD LIIV 540
 Sbjct: 481 SLEDTAKAIEIDTASLIASEDTYVSVTGKGYIKRTSPGFAASTLEEIGKRDDRLIFVQ 540
 Query: 541 NAKTQHLLMFINLGNLMTYRPHHELADIRWQDGEHLSCNLVNFASNEELIYAEVLVDP- 599
 +AKTQHLLMFT LGN+ YRP+HELADIRWQD+GEHLSC + NF +NEEI+Y E+VD F 599
 Sbjct: 541 SAKTQHLLMFTLGNVYIRPHHELADIRWQDIGEHLSCQITNFNEELIYAEVLVDDQFD 600
 Query: 600 TKETYFAVISLQIKRFRQEISPMRTYKSKTAKVAKESVEDVVTVAPIOLEDVIAT 659
 TYFA T LQIKR ER+E +FKRTYKSK+ KYAKLK D +V VAPD+L+DV+L++ 659
 Sbjct: 601 DATTYFAATRLQIKRFRKEFTPMRTYKSKSVKAKLKDDTQIVAVAPIKLDVLLIS 660
 Query: 660 YNGYALRPSINDVPVVGKAGVKAMNLRDRAHIVSAFIANTTSYLLTHERGSLKRAHID 719
 NGYALRF+I +VPVVG+KAGVKAMNLR+ D + SAPI NT+S YLT RGS LKRA+ID 719
 Sbjct: 661 QNGYALRFNTEBVPVVGKAGVKAMNLRKEDDTLQSAFICNTSSFYLLTGRS LKRVSID 720
 Query: 720 VIPTTSRANRGLQVLELRLKPHRVFACPVYLEDSEPFDPSPSVNHEGDTFVLVLEMG 779
 ID TSRA RGLQVLELRLK+KPHRV AG V + F DLPS+ T L+ S 779
 Sbjct: 721 EIPATSRAGRLQVLELRLKPHRVFVLAGSV--AEQGFVGDLPSTEVEENDGT--LLVQS 776
 Query: 780 RTGKVVYVDLSQMSFSEKTSNGSFSDKISDEEVPVKIK 819
 G+Y+ L + SEKTSNGSF+SD ISDEEVP +K 819
 Sbjct: 777 NKGTYTESRLQMLSEKTSNGSFSDTISDEEVPYLYLK 816
 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2823> which encodes the amino acid
 sequence <SEQ ID 2824>. Analysis of this protein sequence reveals the following:
 Possible site: 51
 >> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.53 Transmembrane 376 - 392 (376 - 394)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1213(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
- An alignment of the GAS and GBS proteins is shown below.
- Identities = 633/819 (77%), Positives = 719/819 (87%)
- Query: 1 MSNIGNMSLEIDINSGRFGRYSKYIIQERALPDIRGLKPVQRRILYSNMKDGNTVEKGR 60
 MSNIGNMSLEIDINSGRFGRYSKYIIQERALPDIRGLKPVQRRILYSNMKDGNTVEKGR+R 60
 Sbjct: 3 MSNIGNMSLEIDINSGRFGRYSKYIIQERALPDIRGLKPVQRRILYSNMKDGNTVEKGR 62
 Query: 61 KSAKSGVGNMGNFPHGDSIIYDAMVRMSQDWKNREITLIEHMGKNGSNDGDDPAAMFYTE 120
 KSAKSGVGN+MGNFPHGDSIIYDAMVRMSQDWKNRE L+EMHMGKNGSNDGDD PAAMFYTE 120
 Sbjct: 63 KSAKSGVGNMGNFPHGDSIIYDAMVRMSQDWKNREILVEMHMGKNGSNDGDDPAAMFYTE 122
 Query: 121 ARLSIAGYLLQIDIDKNIVPFAWNPDDTEKEPTVLPAAPFNLVNGATGIGAGYATDIPP 180
 ARLSIAGYLLQDI+KNIV +AWNPDITKKEPTVLPAAPFNLVNG+GIGAGYATDIPP 180
 Sbjct: 123 ARLSIAGYLLQIDIEKNIVSFAWNPDDTEKEPTVLPAAPFNLVNGSGISAGYATDIPP 182
 Query: 181 INLSEVIDAVVYMDIHDPKAKLDMELFLPGDPDFTGIIQGEDEIKKAYETGKRVVRS 240
 INL+SEVIDAVVYMDIHDPKAKLDMELFLPGDPDFTGIIQGEDEIKKAYETGKRVVRS 240
 Sbjct: 183 INLSEVIDAVVYMDIHDPKAKLDMELFLPGDPDFTGIIQGEDEIKKAYETGKRVVRS 242
 Query: 241 RTAETLNGGKKQIIVTETLPEVNEKSVLVRIDDVKNVKNVPGIAEVRDSDRGELRIAI 300

-1020-

- RT IE LKGGK+QIVTEIPYBNK+VLVK+IDVVRNNKVPGI EVRDESDR GLRIAI
 Sb|ct: 243 RTELEKLGKGGQQLIVTNPYBNKAVLVKKIDVVRNNKVPGIVEVRDESDRTGLRIAI 302
- 5 Query: 301 ELKKEADETIVLVNLEFKYTDLVQVNFNFWAIDDYTFKQVGLSRILTSYIAHBEIILAR 360
 ELKKEAD +LNFY KTTDLQVNFNFWAID +TP+QVGL +IL+SYI+HR++IIR R
 Sb|ct: 303 ELKKEADSQTILNLLKCTDLQVNFNFWAIDHPTFRQVGLQKILSSYISHRKDIIER 362
- 10 Query: 361 SKFDKKAERKLHIVEGLIRVLSILDE+IALIR+S+NKADAKENLKVSY+FGF QARAIV 420
 SKFDK KAERKLHIVEGLIRVLSILDE+IALIR+S+NKADAKENLKVSY+FGF QARAIV
 Sb|ct: 363 SKFDKKAERKLHIVEGLIRVLSILDEIITALIRSSUNKADAKENLKVSYDFSEQAIV 422
- 15 Query: 421 TLQLYRLINTDIVTLRESEERLQQITMKAILSDERTMYNVMKRLREVEKKKPANTRYS 480
 TLQLYRLINTDIVTL+ EE +LR IT L AII DE TMYNVHGRELRVKKK PAN R S
 Sb|ct: 423 TLQLYRLINTDIVTLQNSENDLRDLITLSAIGDEATMYNVMHGRELRVKKK PANFRS 482
- 20 Query: 481 ELQELASTIEIDTASLIIEEDTVSVTRGGYVKRISFRSNASTVDLQKRDDELIFVS 540
 ELQ ++ ISIDTASLI EE+T+VSVTRGGY+KRISFRSNAS+++E+GKR+DELIIV
 Sb|ct: 483 ELQELASTIEIDTASLIAEEETFSVTRGGYKRTSFRSNASLEEVRKRDDELIFVK 542
- 25 Query: 541 NAKTTQHLIMFTNLGNLAVFPVHELADIRWQDVEHLNSQNLNFPASNEIITYAEVDPT 600
 AKTT+HLL+FT LGN+ YRP+HEL D+RWKD+QHLISQ + NFA+ BEI+YA++V F
 Sb|ct: 543 QAKTTEHLLFTTLGNVYTRFHELTDLRWQDVEHLISQTTSNFATEEELIYADIVTSFD 602
- 30 Query: 601 KETTYAVTSLQIKRPERQESIPWRTYKSKTAKYKLSVEDYVVVTAPIOLEDLVITY 660
 + Y AVT G IKR+R+E+SPWRTYKSK+ KY KLR +D VVT++P+ +E+L+LT
 Sb|ct: 603 QGLVAVTQNGFIKRFDRKELSPWRTYKSKTKYKLRDDEDRVTLSPVIMEDLLVTK 662
- 35 Query: 661 NGYALRFSINDVPV/GSKAGVYKAMLEKDRDHIVSAFIANTTSLYLTHRGSLKRWAD 720
 NGYALRFS +VP+ G K+AGVK +MLK+ D + SAF + S +LT RGSLEKRWAD
 Sb|ct: 663 NGYALRFSQEVFIQSLKAGVGINLMDSDLSAFAVTSNSPFVLQKRSLEKRWAD 722
- 40 Query: 721 IPTTSRANGLQVLRLELKPFRVFKAGPVLYLSDSFEFDLPSSVSNHEDTFLVLEINSK 780
 IP TSRANGL VLRLEL+KPHRV AG V + S+ +FDLP+ + E + +LE+SK
 Sb|ct: 723 IPQTSRANGLVLRLELTKPHRVFKAGPVLYLSDSFEFDLPDTPDEETNQMLEVISK 782
- 45 Query: 781 TGKTVVDVLSQMSFERTSNGSFVSDKISDEEVSFKIK 819
 TG+ Y++ L S SER GSKSF+SD ISD+EV + +
 Sb|ct: 783 TGQTYIALETLSLSERISNGSFISDTSIDQEVIVAR 821
- 40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 930

A DNA sequence (GBSx0986) was identified in *S. agalactiae* <SEQ ID 2825> which encodes the amino acid sequence <SEQ ID 2826>. Analysis of this protein sequence reveals the following:

- 45 Possible site: 49
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
- 50 bacterial cytoplasm --- Certainty=0.3369 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 55 >GP:AAF64593 GB:AF169649 branched-chain aminotransferase Ilve
 [Lactococcus lactis]
 Identities = 259/340 (76%), Positives = 294/340 (86%)
- Query: 1 MTNLDWNLGFPYKRLPFYIISHFKDKWDGKLTDDATLHISSESPALHYGQCFGL 60
 M +NLDW+NLGF+YK LFFRYI+ FDKGW G+LT D LHISSESPALHYGQ FEG L
 60 Sb|ct: 1 MTNLDWNLGFSYKRLPFYIARFADGWSAGELTODNQLHISSESPALHYGQCFGL 60
- Query: 61 KAYTKDGSIQLEPRDQNAERLQRTADRLIMHPVPTKFLAAKSVVRANEZFPVPGTG 120

-1021-

- KAYRTKDGSIQLFRPDQNA RLQ+TA RL M V T+ FI AVK VV+AN++FVPPYGTG
 Sbjct: 61 KAYRTKDGSIQLFRPDQNAARLQKTARRLCMAEVSSTEMPIDAVKQVVKANKDFVPPYGTG 120
- 5 Query: 121 ATLYIRPLLIGVGDIIIGVKPAEYIPTVFAMPVGSYFKGGLPTINFIVSKEYDRAAPNGT 180
 ATLY+RPLLIGVGD+IGVKPA+EYIF VFAMPVGSYFKGGL P+ F++S+EYDRAAP GT
 Sbjct: 121 ATLYIRPLLIGVGDVIGVKPADEYIFVKFAMPVGSYFKGGLAPSKFVISEYDRAAPLGT 180
- 10 Query: 181 GAAKVGNGYAASLLPGKYAHEKQPSDVIYLDPAHTHTKIEEVGANFFGITKQNFITPLS 240
 G AKVGNGYAASL A ++D IYLDP+THHTKIEEVGANFFGIT DN+FITPLS
 Sbjct: 181 GAAKVGNGYAASLLQAEVGAKASGYADA IYLDPSTHTKIEEVGANFFGITDANEFITPLS 240
- 15 Query: 241 PSILPSITKYSLLYLAKERFGMEAEIGDVFVDELKDFTEAGACGTAAVISPIGGIQND 300
 PSILPSITKYSLLYLA+ R G++AIEG+V+ +L KF EAGACGTAA+ISPIG I +G+D
 Sbjct: 241 PSILPSITKYSLLYLAERHRLGLKAIEGEVYAKDLGKFVEAGACGTAAIISPIGRIDGDD 300
- 20 Query: 301 PHVFSYSETEVGPA+TRKLYDELVGIOQGDVSAPEGMIYKVD 340
 ++F+SETEVGP ++LYDELVGIOQGDVSAPEGMIYKVD
 Sbjct: 301 SYIFHSETEVGPTVKRLYDELVGIOQGDVSAPEGMIYKVD 340
- A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2827> which encodes the amino acid sequence <SEQ ID 2828>. Analysis of this protein sequence reveals the following:
- Possible site: 61
 >>> Seems to have no N-terminal signal sequence
- 25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1208 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
- 30 An alignment of the GAS and GBS proteins is shown below.
- Identities = 280/340 (82%), Positives = 308/340 (90%)
- Query: 1 MT+LNKLNIGFPAYRKLPFRYISHFKDKWDDGKLTDDATLHISESSPALHYGQQAPEGL 60
 MT+ +DKNIGLF Y KLPFRYIS++K+G+ND G+LT+DATLHISES+PALHYGQQAPEGL
 35 Sbjct: 16 MTIAIDNKLNGFETHKLPFRYISYXNGDKWDDGKLTEDATLHISESAPALHYGQQAPEGL 75
- Query: 61 KAYRTKDGSIQLFRPDQNAERLQRTADRLLMPHVPTDKFIAVKSVRANKSFVPPYGTG 120
 KAYRTKDGSIQLFRPD+NA RLQ TADRLLMP V T++FI A K VV+ANE+FVPPYGTG
 Sbjct: 76 KAYRTKDGSIQLFRPDNRNAVRLQATADRLLMPQVSTBQFIDAAKQVVKANEDFVPPYGTG 135
- 40 Query: 121 ATLYIRPLLIGVGDIIIGVKPAEYIPTVFAMPVGSYFKGGLPTINFIVSKEYDRAAPNGT 180
 ATLY+RPLLIGVGDIIIGVKPAEYIPT+FAFVPG+YFKGGL PTINFIVS+ +DRAAP GT
 Sbjct: 136 ATLYIRPLLIGVGDIIIGVKPAEYIPTVFAMPVGNFYFKGGLPTINFIVS+FAFDRAAPGT 195
- 45 Query: 181 GAAKVGNGYAASLLPGKYAHEKQPSDVIYLDPAHTHTKIEEVGANFFGITKQNFITPLS 240
 GAAKVGNGYA SLLPGK A PSDVIYLDPAHTHTKIEEVGANFFGIT +N+F+TPLS
 Sbjct: 196 GAAKVGNGYAASLLPGKAASAGSPSDVIYLDPAHTHTKIEEVGANFFGITANNEFVTPLS 255
- 50 Query: 241 PSILPSITKYSLLYLAKERFGMEAEIGDVFVDELKDFTEAGACGTAAVISPIGGIQND 300
 PSILPSITKYSLLY LA+ER GM IEEDV ++ELDKF EAGACGTAAVISPIGGIQ D+
 Sbjct: 256 PSILPSITKYSLLYLAERLGMTVIEGIVFVIELDKFVEAGACGTAAVISPIGGIQYKND 315
- Query: 301 PHVFSYSETEVGPA+TRKLYDELVGIOQGDVSAPEGMIYKVD 340
 HVFSYSETEVGP TR+LYDELVGIOQGD+SAPEGMIYKVD
 55 Sbjct: 316 LHVFSYSETEVGPTRRLYDELVGIOQGDVSAPEGMIYKVD 355

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 931

- 60 A DNA sequence (GBSx0987) was identified in *S.agalactiae* <SEQ ID 2829> which encodes the amino acid sequence <SEQ ID 2830>. Analysis of this protein sequence reveals the following:

-1022-

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

5 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3459(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9365> which encodes amino acid sequence <SEQ ID 9366>
 10 was also identified. A further related GBS nucleic acid sequence <SEQ ID 10915> which encodes amino acid sequence <SEQ ID 10916> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2831> which encodes the amino acid sequence <SEQ ID 2832>. Analysis of this protein sequence reveals the following:

15 Possible site: 50
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3043(Affirmative) < succ>
 20 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

25 Identities = 22/36 (61%), Positives = 30/36 (83%)
 Query: 4 IVSKDKKIEIQISDAQVTVNGTKVDGYQLVMEKKL 39
 ++SKDKKIEIQ+ D +V VN TK+DGYQL + K++
 Sbjct: 1 VMSKKDKKIEQLIDHKVMVNETKIDGYQLQIGKRV 36

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 932

A DNA sequence (GBSx0988) was identified in *S.agalactiae* <SEQ ID 2833> which encodes the amino acid sequence <SEQ ID 2834>. This protein is predicted to be glycyl-tRNA synthetase beta subunit (glyS).

35 Analysis of this protein sequence reveals the following:

 Possible site: 14
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 40 bacterial cytoplasm --- Certainty=0.1617(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:CAE73488 GB:AL139077 glycyl-tRNA synthetase beta chain
 [Campylobacter jejuni]
 Identities = 33/90 (36%), Positives = 49/90 (53%), Gaps = 2/90 (2%)
 Query: 3 RAFNLAERKVTHTSVLDSSLPENKQKALYQAILSLLETDMHNLKLFALPSIINDFPD 62
 R N+A K H V D SLF E LY+A + + L+ LEAL P I++FF+
 50 Sbjct: 570 RLANIATKNPHKV--DESLFVQASRSLKYKAFQRKTSNLQKRLNLFALPKPFIDFPN 627
 Query: 63 NTMVMITDDEKMKQRLIALINSLVAKARTVA 92
 M+ +DEK+K NR A++ + A+ +A

-1023-

Sbjct: 628 QVMINAEDKELKNNRQALVYRIYAEPLKIA 657

There is also homology to SEQ ID 2836.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 933

A DNA sequence (GBSx0989) was identified in *S.agalactiae* <SEQ ID 2837> which encodes the amino acid sequence <SEQ ID 2838>. Analysis of this protein sequence reveals the following:

Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.4825 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13672 GB:Z99113 ynzC [Bacillus subtilis]
Identities = 41/72 (56%), Positives = 56/72 (76%)

Query: 5 KIARINELSKKKKTVGLTGEEKVQAKLRREYIEGPRRSVRHHVGEIKLVDDGNDVTE 64
KIARINEL+ K K +T EEK EQ KLR+EY++GFR S+++ ++ +K++D EGNVDVTE
Sbjct: 6 KIARINELAAKAKAGVITREEKAEQQKLRQEYLKGFSSMQNTLKSVKLIDPEGNDVTE 65

Query: 65 KLRQVQRKGLH 76
KL++ QR LH
Sbjct: 66 KLRQVQRKGLH 77

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2839> which encodes the amino acid sequence <SEQ ID 2840>. Analysis of this protein sequence reveals the following:

Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.4303 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 79/85 (92%), Positives = 83/85 (96%)

Query: 1 MDPKKIARINELSKKKKTVGLTGEEKVQAKLRREYIEGPRRSVRHHVGEIKLVDDGND 60
MDPKIARINEL+KKKKTVGLTG EKVQAKLRREYIEG+RRSVRHH+BGTKLVD+EGND
Sbjct: 1 MDPKKIARINELAKKKKTVGLTGEEKVQAKLRREYIEGYRRSVRHH+BGTKLVD+EGND 60

Query: 61 VTPEKLRQVQRKGLHGRSLDDPNS 85
VTPEKLRQVQRKGLHGRSLDDP 8
Sbjct: 61 VTPEKLRQVQRKGLHGRSLDDPKS 85

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 934

A DNA sequence (GBSx0990) was identified in *S. agalactiae* <SEQ ID 2841> which encodes the amino acid sequence <SEQ ID 2842>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2343 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AA669985 GB:U94355 glycerol kinase [Enterococcus casseliflavus]
Identities = 381/496 (76%), Positives = 439/496 (87%)

Query: 3 SEEKYIMALDQGITSSRAIIFNKKGEKIASQKEFPQIFPQAGMVEHNANQIWNVSQSVI 62
+E+ Y+MALDQGITSSRAIIF+ G+KI SSQKEFPQ FP+GMVEHNAN+IWNVSQSVI
Sbjct: 2 AEQYIVMALDQGITSSRAIIFERNKKIGSSQKEFPQYFPKSGMVEHNANEIWNVSQSVI 61

Query: 63 AGAFIESIKPQOIEAIGITNQRETTVVMDKKITGLPIYMAIVMQSRQTAPIADQLKQEGH 122
AGAFIES I+P I IGITNQRETTVVMDK TG PI NAIVMQSRQ++PIADQLK +GH
Sbjct: 62 AGAFIESGIRPBAIAGITNQRETTVVMDKITGCPINAIIVMQSRQSSPIADQLKVDGH 121

Query: 123 TMIHEKTKGLVIDAYFSATKVRMILDRHVPGAQERAKGELLFGTIDTWLVKILTDGLVHV 182
T MIHEKTKGLVIDAYFSATKVRM+LD++ GAQE+A+ GELLFGTID+MLVVKILTDG VHV
Sbjct: 122 TMIHEKTKGLVIDAYFSATKVRMILDNIBGAQEKADNGELLFGTIDSLVVKILTDGQVHV 181

Query: 183 TDYSNAARTMLYNIKELQNDREILBLLNIPKAMLPVEKSNSEVYQKTIPTFFHYGGSEVPS 242
TDYSNA+RTMLYNI +L+ND EIL+LLNIP +MLPVEKSNSEVYQ T +HFYQ EVI+
Sbjct: 182 TDYSNASRTMLYNIHKLEWDREILDLNIPSSMLPVEKSNSEVYQKTSRHYGSEVPS 241

Query: 243 GWAGDQQAALFQQLAFEPGMVQNTYGTGSFIIMTGEEMQLSQNNLLTTIGYGINGKVHY 302
GWAGDQQAALFQQL+AFE G+KNTYGTG+FI+MTTGE QLS N+LTTTIGYGINGKV+Y
Sbjct: 242 GWAGDQQAALFQQAQAFKGMKIKNTYGTGAFIVMTNTEEPQLSDNDLTTTIGYGINGKVYY 301

Query: 303 ALBGSIFIAGSAIQWLRDGLRMIETSSSEBGLAQSTSDDEVIYVVPFTGLGAPYVDSNA 362
ALBGSIF+AGSAIQWLRDGLRMIETSS+SE LA + D+EVYVVPFTGLGAPYVDS A
Sbjct: 302 ALBGSIFVAGSAIQWLRDGLRMIETSPQSEBLAAKAGKNEVYVVPFTGLGAPYVDSRA 361

Query: 363 RGSVFGILTRGTSKEDFVKATQLSIAYQVRVDITMQVDSGIDIQQLRVDGGAAENLLMQ 422
RG+VFGILTRGT+KEDFV+ATLQ++AYQ +DVIDTM+ DSGIDI L+VDGGAA N+LLMQ
Sbjct: 362 RGSVFGILTRGTTKEDFVRATLQAVAYQSKDVIDITMKDSGIDIPLLKVDGGAAENLLMQ 421

Query: 423 FQADILIGIDIAKAKNLETTALGAAFLAGLSVGYWESMDELKELNATGOLFQATYNESRKE 482
FQADIL ID+ RA NLETTALGNA+LNL+VG+W+ +DELK + GQ+F M ++
Sbjct: 422 FQADILIDVORANLETTALGAAYLAGLAVGFWKIDDELKSMAREEQNTMTPMAEERD 481

Query: 483 KLYGWRKAVKATQVF 498
LY+G+H+AV ATQ F
Sbjct: 482 NLYGWRKQAVATQT 497

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2843> which encodes the amino acid sequence <SEQ ID 2844>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2282 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

-1025-

Identities = 464/500 (92%), Positives = 484/500 (96%)

Query: 3 SERKYIMAIIDQGTSSRAIIPNKKGGKLIASSQKEFPQIPFQAGNVEHNAQIWNISVQSVI 62
 5 S+EKYIMAIIDQGTSSRAIIPN+KGEK+SSQKEFPQIPF AGNVEHNAQIWNISVQSVI
 Sbjct: 2 SQEKYIMAIIDQGTSSRAIIPNQGKEKVSSSQKEFPQIPFAGNVEHNAQIWNISVQSVI 61

Query: 63 AGAPTESSIKPQIEAIGITNQRETTVVWDKKTGLPIYNAIVWQSRGTAPIADQLKQGH 122
 AGAPTESSIKP QIEAIGITNQRETTVVWDKKTG+PIYNAIVWQSRGTAPIA+QLKQ+GH
 10 Sbjct: 62 AGAPTESSIKPQIEAIGITNQRETTVVWDKKTGVPYNAIVWQSRGTAPIADQLKQGH 121

Query: 123 TMIHEKTGLVIDAYFSATKIRWILDHVPGAQERAEKGEELPGTIDTMDLWKLTDGLVHV 182
 T MIHEKTGLVIDAYFSATKIRWILDHVPGAQERAEKGEELPGTIDTMDLWKLTDG VHV
 Sbjct: 122 TMIHEKTGLVIDAYFSATKIRWILDHVPGAQERAEKGEELPGTIDTMDLWKLTDGAVHV 181

Query: 183 TDYNSAARTMLYNIKELWDEILELLNIPKAMLPFKVKSNSFVYGTTPHFYGGVFPIS 242
 TDYNSAARTMLYNIK-L WDEILELLNIPK MIFVKNSSE+YGT FHFYGGVFPIS
 15 Sbjct: 182 TDYNSAARTMLYNIKDLTWDEILELLNIPKMLPFVKNSSEIYGTAPHFYGGVFPIS 241

Query: 243 GWAGDQQAALFQQLAFEPGFWQNTYGTGSLIIMVTEGEMQLSQNNLLTTIGYINGKVHY 302
 GWAGDQQAALFQQLAFEPGFWQNTYGTGSLIIMVTEGEMQLSQNNLLTTIGYINGKVHY
 20 Sbjct: 242 GWAGDQQAALFQQLAFEPGFWQNTYGTGSLIIMVTEGEMQLSQNNLLTTIGYINGKVHY 301

Query: 303 ALEGSIFINGSAIQWLRLDGLMIETSSSEGLAQSTSDDEVVVVPAFTGLGAPYWDNA 362
 ALEGSIFINGSAIQWLRLDGLMIETSSSEGLAQSTSDDEVVVVPAFTGLGAPYWDNA
 25 Sbjct: 302 ALEGSIFINGSAIQWLRLDGLMIETSSSEGLAQSTSDDEVVVVPAFTGLGAPYWDNA 361

Query: 363 RGSVFLTRGTSKEDFVKATLQSIAYQVRDVIDTMQVDSGIDIQLLRVDGGAMNNMLAQ 422
 RGSVFLTRGTSKEDFVKATLQSIAYQVRDVIDTMQVDSGIDIQLLRVDGGAMNNMLAQ
 30 Sbjct: 362 RGSVFLTRGTSKEDFVKATLQSIAYQVRDVIDTMQVDSGIDIQLLRVDGGAMNNMLAQ 421

Query: 423 PQADILGIDIRAKNLETTALGAFLAGLGVYWE MD LKEINATGQLF+A+MNSRKE 482
 PQADILGIDIRAKNLETTALGAFLAGLGVYWE MD LKEINATGQLF+A+MNSRKE
 Sbjct: 422 PQADILGIDIRAKNLETTALGAFLAGLGVYWE MD LKEINATGQLF+A+MNSRKE 481

Query: 483 KLYGKRAKAVKATQVFAQED 502
 KLYGKRAKAVKATQVF QE+
 35 Sbjct: 482 KLYGKRAKAVKATQVFAQED 501

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 935

A DNA sequence (GBSx0992) was identified in *S. agalactiae* <SEQ ID 2845> which encodes the amino acid sequence <SEQ ID 2846>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3146 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 50 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 936

A DNA sequence (GBSx0993) was identified in *S. agalactiae* <SEQ ID 2847> which encodes the amino acid sequence <SEQ ID 2848>. This protein is predicted to be alpha-glycerophosphate oxidase (glpD).

Analysis of this protein sequence reveals the following:

```

5      Possible site: 40
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.81    Transmembrane    20 - 36 ( 20 - 36)

10      ----- Final Results -----
      bacterial membrane --- Certainty=0.1723 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15      >GP:AC34740 GB:U94770 alpha-glycerophosphate oxidase [Streptococcus pneumoniae]
      Identities = 464/608 (76%)    Positives = 539/608 (88%)

Query: 1      MEFSRETRRLALQRMQDRTLDLLI IGGGITGAGVALQAAASGLDTGLIEMQDFAESTSSR 60
MEFS++TR L+++++MQ+RTLDLLI IGGGITGAGVALQAAASGL+TGLIEMQDFAESTSSR
20      Sbjct: 1      MEFSKTRRELSTIKMQDERTLDLLI IGGGITGAGVALQAAASGLDTGLIEMQDFAESTSSR 60

Query: 61      STKLVRHGLRLYLKQFDVEVSDTVSRAVVQQTAPHIPKPDMLLPVYDEPGSTFSMFLR 120
STKLVRHGLRLYLKQFDVEVSDTVSRAVVQQTAPHIPKPDMLLPVYDE G+TFS+FLR
20      Sbjct: 61      STKLVRHGLRLYLKQFDVEVSDTVSRAVVQQTAPHIPKPDMLLPVYDEGATFSFLR 120

Query: 122     KVAMLDYLLAGVNTPTANKVLSARDVLRKFPDLQKEGLLGGGVYLDPRNDARLVLEN 180
KVAMLDYLLAGV+NTPTANKVLS +VL+R+P+L+KEGL+GGGVYLDPRNDARLVLEN
25      Sbjct: 122     KVAMLDYLLAGVNTPTANKVLSKDQVLRQPNLKKGLVGGGVYLDPRNDARLVLEN 180

Query: 181     IKRANRDGAY IASHVKAEDFLFDNNQIIGVRARDLLTDQVIDI KARLVINTTGPWSDTV 240
IKRAN+DGA IA+HVKA EFLFD++ +I GV ARDLLTDQV +IKARLVINTTGPWSDTV
30      Sbjct: 181     IKRANQCGALLANHVKAEGFLFDSGKITGVVARDLLTDQVFEI KARLVINTTGPWSDTV 240

Query: 241     RNFSNBEKQIHQLRPTKGVHLVDRLQKINISQPVVYDTGLNDGRMI FVLPRDKTYFGTT 300
RN SN+G Q Q+RPTKGVHLVDV K+ +SQPVY DTGL DGRM+FVLPR+KTYFGTT
35      Sbjct: 241     RNLSNKGTPQSPQMRPTKGVHLVDSKIKVSCPVTFTDGLDGRMVFLPRENKTYFGTT 300

Query: 301     DTDYHGDELHPVTYKEDVDYLLINIVNRPFEAEITDIDESSWAGLRPLLSGNSASDYNG 360
DTPY GDELHP VT+EDVDYLL IVN RPE+ +TIDIDESSWAGLRPL++GNSASDYNG
40      Sbjct: 301     DTDYHGDELHPVTYKEDVDYLLIGLVNRPFEENITDIDESSWAGLRPLLSGNSASDYNG 360

Query: 361     GNSGLSDSPFEELIDSVKDYIAHKNHREDVEKAISHVESSTSEKELDPSAVSRGSSPER 420
GN+G +SDSPF+ LI +V+ Y++ + RDEVE A+S +ESTSEK LDPASVSRGSS +R
45      Sbjct: 361     GNGGTSDESPFNLIATVESYLSKEXTREDVESAVSKLESSTSEKILDPASVSRGSSLD 420

Query: 421     DNGLLTLAGGKITDYRKMAEGAMEITIINILDKETHRKPKLINSKITYPVSGGEINPNDV 480
DNGLLTLAGGKITDYRKMAEGAME +++IL E++R PKLINSKITYPVSGGE+NP+NDV
50      Sbjct: 421     DNGLLTLAGGKITDYRKMAEGAMEKVVILKAEFDRSPKLINSKITYPVSGGEINPNDV 480

Query: 481     SEIEHAYQLGTLSGLSIEDARYIANLYGNSAPKFAIATRQITSEAGLSIVETLSLHAYMD 540
SEIEH+AQIG GL ++A Y+ANLYGNSAPK+PAL ++A GLSL+TSLSHYAM
50      Sbjct: 481     SEIEHAFQLGVSRGLDSKEAHYIANLYGNSAPKVFALAHSLQCAPGLSLADTSLSHYAM 540

Query: 541     YEMALSPTDFLRKTNHMLFMRDLSLSIQPVIDEMAKHYQMSDQKTPYEEELHSTLKD 600
E+ALSP DF LKRTNHMLFMRD+LDS+++PV+DEM +Y H++++K Y++ L+
55      Sbjct: 541     NELALSPTDFLRKTNHMLFMRDLSLSIVKPVLDGRGRFYDWTREKATYRADVRAALAN 600

Query: 601     NDLAALAD 608
NDLA LK+
60      Sbjct: 601     NDLAELAN 608

```

There is also homology to SEQ ID 128.

-1027-

SEQ ID 2848 (GBS93) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 7; MW 70.6kDa).

GBS93-His was purified as shown in Figure 192, lane 4.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 937

A DNA sequence (GBSx0994) was identified in *S.agalactiae* <SEQ ID 2849> which encodes the amino acid sequence <SEQ ID 2850>. Analysis of this protein sequence reveals the following:

```

Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0965(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 938

A DNA sequence (GBSx0995) was identified in *S.agalactiae* <SEQ ID 2851> which encodes the amino acid sequence <SEQ ID 2852>. This protein is predicted to be glycerol uptake facilitator protein (glpF). Analysis of this protein sequence reveals the following:

```

Possible site: 55
>>> Seems to have a cleavable N-term signal seq.
INTEGRAL Likelihood = -7.43 Transmembrane 220 - 236 ( 216 - 236)
INTEGRAL Likelihood = -6.48 Transmembrane 139 - 155 ( 136 - 158)
INTEGRAL Likelihood = -3.88 Transmembrane 87 - 103 ( 83 - 107)
INTEGRAL Likelihood = -3.03 Transmembrane 164 - 180 ( 162 - 183)

----- Final Results -----
bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8689> which encodes amino acid sequence <SEQ ID 8690> was also identified. Analysis of this protein sequence reveals the following:

```

Lipcp: Possible site: -1 Crend: 5
SRCFLG: 0
McG: Length of UR: 21
Peak Value of UR: 2.51
Net Charge of CR: -2
McG: Discrim Score: 4.43
GvH: Signal Score (-7.5): -0.139999
Possible site: 50
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 51
ALOM program count: 4 value: -7.43 threshold: 0.0
INTEGRAL Likelihood = -7.43 Transmembrane 215 - 231 ( 211 - 231)

```

-1028-

INTEGRAL Likelihood = -6.48 Transmembrane 134 - 150 (131 - 153)
 INTEGRAL Likelihood = -3.88 Transmembrane 82 - 98 (78 - 102)
 INTEGRAL Likelihood = -3.03 Transmembrane 159 - 175 (157 - 178)
 PERIPHERAL Likelihood = 4.98 65
 5 modified ALOM score: 1.99
 icm1 HYPID: 7 CFP: 0.397
 *** Reasoning Step: 3
 10 ----- Final Results -----
 bacterial membrane --- Certainty=0.3972 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
 15 The protein has homology with the following sequences in the GENPEPT database.
 >GP:AA91618 GB:U12567 glycerol uptake facilitator [Streptococcus pneumoniae]
 Identities = 150/230 (65%), Positives = 194/230 (84%), Gaps = 1/230 (0%)
 20 Query: 7 DIPGEFLGTALLVLLGNVAGVVLPEKTHASGNIVITFGWGLAVAVAVFVSGKVPAAH 66
 ++PGEFLGT +L+LLGNVAGVVLPEK++SGWIVIT G+AVA+A VSG+SPAH
 Sbjct: 4 ELPGELGTLLILLGNVAGVVLPEKTHSSSGWIVITW+GIAVAVAVFVSGKSPAH 62
 Query: 67 LNPVSLAFANGGLAWGTALLVLLGNVAGVVLPEKTHASGNIVITFGWGLAVAVAVFVSGKVPAAH 126
 LNPV++ A+KG L W++Y+AQ GAMLG +LV+LQF+PHYEA EN +IL TP+
 25 Sbjct: 63 LNPVTTIGVALLGGLPWASVLPYLLAQFAGAMLGQILVWLQFPHYEAENAGHILATFS 122
 Query: 127 TGPALCNFNSFLSEVLGTLLVLLTIFAIKGNMPFGVGTMSVGLVVGIGLSGGTGTG 186
 TGP+KD SN +SE+LGT VLVLTIFA+G Y+ G+GT +VG L+VGIGLSGGTGTG
 Sbjct: 123 TGPALKDVSNLSEILGTFLVLLTIFALGLVDFQAGIGTFAVGLTIGIGLSGGTGTG 182
 30 Query: 187 ALNPARDFGPRLHALLFMKNGKSDMTYWNIPVGNVGAIALIAPAM 236
 A+NPARD GPR++H++LE+ NKGD DM+Y+WIP+VGP++GA LA L+V++
 Sbjct: 183 ALNPARDLGPRLMSILFIPKNGKSDMSYANIPVGVGTGAALAVLVSIL 232

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2853> which encodes the amino acid sequence <SEQ ID 2854>. Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have a cleavable N-term signal seq.
 40 INTEGRAL Likelihood = -9.13 Transmembrane 213 - 229 (209 - 232)
 INTEGRAL Likelihood = -5.52 Transmembrane 137 - 153 (132 - 157)
 INTEGRAL Likelihood = -4.35 Transmembrane 159 - 175 (155 - 178)
 INTEGRAL Likelihood = -1.17 Transmembrane 85 - 101 (85 - 101)
 45 ----- Final Results -----
 bacterial membrane --- Certainty=0.4652 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

50 >GP:AA91618 GB:U12567 glycerol uptake facilitator [Streptococcus pneumoniae]
 Identities = 159/230 (69%), Positives = 196/230 (85%), Gaps = 1/230 (0%)
 Query: 2 DIPGEFLGTALLVLLGNVAGVVLPEKTHASGNIVITFGWGLAVAVAVFVSGKVPAAH 61
 ++PGEFLGT +L+LLGNVAGVVLPEK++SGWIVIT T GIAVAVAVFVSGK+PAH
 55 Sbjct: 4 ELPGELGTLLILLGNVAGVVLPEKTHSSSGWIVITW+GIAVAVAVFVSGKSPAH 62
 Query: 62 LNPVSLAFANGGLAWGTALLVLLGNVAGVVLPEKTHASGNIVITFGWGLAVAVAVFVSGKVPAAH 121
 LNPV++ A+G+W++Y LAQ GAM+G LV+LQF+PHY A E+ +IL TP+
 60 Sbjct: 63 LNPVTTIGVALLGGLPWASVLPYLLAQFAGAMLGQILVWLQFPHYEAENAGHILATFS 122
 Query: 127 TGPALCNFNSFLSEVLGTLLVLLTIFAIKGNMPFGVGTMSVGLVVGIGLSGGTGTG 186
 TGPAL+DT SNL+SEI GTFLV+L I A GLYD AG+GT VGTL+GIGLSGGTGTG
 Sbjct: 123 TGPALKDVSNLSEILGTFLVLLTIFALGLVDFQAGIGTFAVGLTIGIGLSGGTGTG 182

-1029-

Query: 182 AINPARDLGPRILVHAILPANNKGDSDWSYANIPVVGPIIGAVLAVLLFQV 231
 A+NFPARDLGPR+H+L+P+ NKGD DMSYANIPVVG+TGA LAVL+P +
 Sbjct: 183 ALNPARDLGPRIMHSILEPINKGDGDSYANIPVVGPIIGAVLAVLVSFL 232

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 169/232 (72%), Positives = 202/232 (86%)

Query: 6 MDIFGSEFLGTALLVLLGKGVAGVVLPTKKNHNSCWIVITFGWGLAVAIALVSGNISPA 65
 MDIFGSEFLGTALLVLLGKGVAGVVLPTK H SSWIVI GWG+AVA+A +SG ++PA
 10 Sbjct: 1 MDIFGSEFLGTALLVLLGKGVAGVVLPTKTHASSWIVATGWLAVAVAFISGKVAFA 60

Query: 66 HLNPAVSLAFAIKGDLAWGTAILYMIQIIGMLGSLVYLQFRPHYEAENRADILGTF 125
 HLNPAVSLAFA+ G +AW TAI Y +AQ++GMA+GS LV+LQFRPHY AAE++ADILGTF
 15 Sbjct: 61 HLNPAVSLAFAMSGTIAWSTALAYSIAQLLGMVGSILVFLQFRPHYLAESQADILGTF 120

Query: 126 ATGPALKDNFSNFLSSVLSTLVLLVLTIFAIGKKNMFGVGTMSVGMVLVVGIGLSLGGTIG 185
 ATGFA+D SN LSE+ GT VL+L I A G Y+MP G+G+ VG LV+GIGLSLGGTIG
 15 Sbjct: 121 ATGPARTDSSNLLSSIPGTFVLMGILAFGLYDMFAGLGLTLCVGTLVIGIGLSLGGTIG 180

Query: 186 YAINPARDLGPRLLHALLPMNKGDSDWYISWIPVGMVGAIALAALIPAM 237
 YAINPARD GPRL+HA+L+P+ NKGDSDW+Y+WIP+VGP++GA+LA L+P +M
 20 Sbjct: 181 YAINPARDLGPRILVHAILPANNKGDSDWSYANIPVVGPIIGAVLAVLLFQV 232

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 25 vaccines or diagnostics.

Example 939

A DNA sequence (GBSx0996) was identified in *S. agalactiae* <SEQ ID 2855> which encodes the amino acid sequence <SEQ ID 2856>. This protein is predicted to be NADH oxidase. Analysis of this protein sequence reveals the following:

30 Possible site: 23
 >> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -2.87 Transmembrane 152 - 168 (152 - 168)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.2147(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9523> which encodes amino acid sequence <SEQ ID 9524>
 40 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA48728 GB:X68847 NADH oxidase [Enterococcus faecalis]
 Identities = 105/423 (24%), Positives = 197/423 (45%), Gaps = 15/423 (3%)

45 Query: 10 IVILGASFAGMTCAQKLRQINPNNDIVLDKEITHPYDINGLNWYRHEISGLAQPMWOT 69
 +V++G +AG + + + +P ++ + + + +G+ Y + +
 Sbjct: 3 VVVVGCTHAGTSAVKSLIANHPEAVTVYKRNNDISFLSGTIALYVGGVVKNAALFYSN 62

50 Query: 70 EEEQRLQNRCLFGLKVEKINKEDR-----ELMLSDGSSVYVDQLTCAMSGQABSTYIDG 124
 EE VE+IN +D+ L +V YD+L+ GS I G
 Sbjct: 63 PEELASLGATVYKGBNVEINVDKTKVIAKLQATGTETVSYKLVMTTGSMPITPPPG 122

Query: 125 ADAQGVLTTKTYATSCNAQKVLQDKSHKVAVVGAGIIGLDIAYSILHSGKAVTLEAQERP 184
 DA+ +L K Y+ + + + +V VVG G IG++ + + BSGK VTI++ +R
 55 Sbjct: 123 IDAENILLCKNYSQANVTIEKAKDAKRVVVVGGYIGIELVEAFVBSGKQVTLVDGLDR 182

Query: 185 DFRHTDPMSLPILDAMAEKSLHFFONQKVEKITVTRKKLCRLRTGTDTFTVDVILAV 244
 ++ D + L + + + + + V++ + K+ F D VI + V

-1030-

Sbjct: 183 INKYLDKPPTDVELEKELVDRGVNLALGENVQQFVADBQKVKVITPQSQFEADVMHCV 242
 Query: 245 NFRPDRRLTGLVDLSVNSVVDYPTSDPNTYAIDLINSYPKGLNSAYMPLINQA 304
 FRP++ LL VD+ ++ VM+Y QTS+P+I+A GD ++ + Y+EL A
 Sbjct: 243 GFRINTELLKDKVDMLPKRAIEVNEYMTQTSNFDI PAAGDSAVVHYNPSQTKRYIPLATNA 302
 Query: 305 IRSQMIAYHLGSHVAPKLKITRATGSKHFPYRANIGLT-----ELEAGPDETIV 355
 +R ++ +L+ + +G PG+ +G+T ++EA +ED
 Sbjct: 303 VHQGLMGRNLTETQLAVRGTQGTSGLYLPKWKIGSGVYKESAKRLNGLDEATVPELRY 362
 Query: 356 SVTYFPKQYDLRIKLIANKQTHILGAQLISKENCLATNMQLVQAISCDWTDDELAFQD 415
 + P + L ++L+ + T ++G QGL+SK + +M L A + MT DLA D
 Sbjct: 363 RPRFMPPTKVL-MELVVEKGTQRIVGGGLMSKYDITQSAVHLSLAVQNMVTEDELALSD 421
 Query: 416 FIY 418
 F +
 Sbjct: 422 FFF 424

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2857> which encodes the amino acid
 sequence <SEQ ID 2858>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -3.35 Transmembrane 155 - 171 (155 - 173)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2338 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
 RGD motif: 54-56

The protein has homology with the following sequences in the databases:

>Q:CAA44611 GB:X62755 NADH peroxidase [Enterococcus faecalis]
 Identities = 111/428 (25%), Positives = 202/428 (46%), Gaps = 24/428 (5%)
 Query: 10 VICASFAGLAFVDKYLKDLNFDSDQIILLIKESCPNYIHWJINQLFRGDIQDLSDMNGRAC 69
 V+G+S G V++ +L+ED++I +K +++ G+ G++D++ R
 Sbjct: 5 VLGSSHGQYEAVERELNLHFDATQWYHGDPIISFLACGMQLVLEGGKVDVNSV---RYM 61
 Query: 70 LAAQIESN--HRFQAQVLAIEAPSNITLLKDS-QGRVFESEGVETLVCMAGASQGHYIE 126
 ++ES + F E+ A++ ++ +KD G E Y+ L+ + GA P I
 Sbjct: 62 TGEIOMESRGVNVFSNTEITAIQPKHQVTVKDLVSGEERVENYDKLTIISGAVPPELDIP 121
 Query: 127 TSGTNKVLVTKYTESQASLKLIEASOR-----VLVIGAGLGLDLAYSLSLQGRKRVLI 181
 + + + + Q ++KL + + + V+VIG+G IGA+ A + + GK+V +I
 Sbjct: 122 GKDLNITLYMR---GRQNAIKLKQXTVDPEVNVVVIGSGYIGIEAAEAPAKGKGVTVI 178
 Query: 182 EAAERPDFYQTDARLAPVMAEMSTHVTFINNKRVTAIHKLIRGKVVAHTEQGDTPGDDL 241
 + +RP D E + EM +++T + V +E +G+V ++ DL
 Sbjct: 179 DILRFLGVYLDKEPTDVLTEMEANITITATGETVER-YKDGGRVQKVPVIDGAXDADL 237
 Query: 242 AILAINRFRNTHLQCCVACALDKITLWENLQTSQANIYAICDMVSLHFGILGMGYTTP 301
 ++A+ RNT L+G + + I +E ++TS+ ++A+GD + + +
 Sbjct: 238 VVAVGVRENTANLKSTLELHNGLIKTDYEMRTSEPDVFAVGQATLIKYNPAETVENIA 297
 Query: 302 LINQAMKIQGALALHLAGYPIPLQTVK-VLGSSHPDYRASGVTE-----EAEELY 353
 L A K G+ +L P+ P V+ G +FDY AS G+ E +E+
 Sbjct: 298 LATNARKQGRFAVKNLE-EPVKPPFGVQSSGLAVFDYKPASTGINEVMAQKLKGTAKV 356
 Query: 354 MDTCSYLYQNGSDNRLFKLIRKIDGILLIGALSKTNALVIANQLGQALAKVITAD 413
 YL K W KL+ ++GAQL+SK + N + A+ K+T D
 Sbjct: 357 TVVEDYLDMPNFDKQKAMPFLVYDFETTLQGLAQMSKADLTANINAIISLIQAKMTIED 416
 Query: 414 IAPQDFLF 421

-1031-

LH+ DF F

Sbjct: 417 LAYADFFF 424

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 192/440 (43%), Positives = 276/440 (62%), Gaps = 7/440 (1%)

Query: 8 KIVIVLGASFAQMTCAQLRQLNPNWDIVLIDKEIHDPYVPGNLNMYRHEISGLNQAMW 67
 K I ++GASFAG+ K + LNP+ I-LIDKE P+Y+PNG+N +R +I L+ AAM
 10 Sbjct: 6 KTHIVIGASFAGLAFVDKYKDLNPSQIILLIDKESCFYITPNGINQLPRGDIQDLSDAMW 65

Query: 68 -QTEEBQRLQIRNIRCLFGKVKVKINKEDRELMLSDSSVY----YDQLICAMGSAASYI 122
 + ++++ +V I L+L D Y+ L+CMG+ +S YI
 15 Sbjct: 66 GRACLAQAQIRSNHRFTQAQEVLAIRAPSNITLLKDSQGRVFEQGYETLVCAMGASQSHYI 125

Query: 123 DGADAQSVITTKTYATSNNAKQVLDRSHKVAVVGAGIIGLDIAYSLSHSGKAVTLEAQE 182
 + + VL TK Y SQ + +++ S +V V+GAG+IGLD+AYS L GK V L+EA E
 20 Sbjct: 126 BTSQNKVNLVTKYESSQASLKITRASQEVAVGAGLIGLDLAYSLSLQKRVKILRAAF 185

Query: 183 RPDFRHTDPDMSLPLLDAMARESKLHFFQNKQVEKITVTRBKLCIKRTLTGDTFTVDAIL 242
 RPDF TD ++ P++ M+ + F N++V I E K+ T GUTP D IL
 25 Sbjct: 186 RPDFYQTDABLIAFVMAEMSTHRVTF INNKRVTAIHEI-BGKVVAHTEQGDFTQGDLAAIL 244

Query: 243 AVNFRPDSRLTLGLVDSVNSVVNDYFQTSDPNIYAIGDLWISYFKGLNSAYYPLIN 302
 A+NFPP+ LL G V ++D ++VM+ QTS NIYAIGD++ +F L YI PLIN
 30 Sbjct: 245 AINFPPNYHLLGQVACALDKTILVNMENLQTSQANIYAIGDMVSLHFGILGMDYYPLIN 304

Query: 303 QAIRSAQMLAYLHSGHVPFLKLTIRATSGSKHFGYVRANIGLLEAGFYEDTVSVITYPK 362
 QA+++ Q LA HL+G+ +P L+ + GS HP YYRA++G+TE EA Y DF S Y
 35 Sbjct: 305 QAMKTGQALALHLAGTYPIPELQTVRVLGGSHDFYTRASVGVTEEBEARLYMDTCSYLYQNG 364

Query: 363 BQYDL-RIKLIANKTGHLLGAQLISKENCILATANQLVQAISCDNTDFDLAQDFPIYAR 421
 + +L +KLIA + G L+GQL+SK N L ANQL QA++ +TD DLAQDF+
 40 Sbjct: 365 DSNKLFWLKLARKTDGILIGAGQLSKTNALVIANQLGQALALKVTDADLAQDFLFG 424

35 Query: 422 ESEMAVMLHQAALNLYEKRI 441
 S++AY LH+A + L+ERR+
 Sbjct: 425 HSDLAYHLEACKLFEKRL 444

There is also homology to SEQ IDs 1820, 1876, 4666.

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 940

A DNA sequence (GBS0998) was identified in *S.galactiae* <SEQ ID 2859> which encodes the amino acid sequence <SEQ ID 2860>. Analysis of this protein sequence reveals the following:

45 Possible site: 31
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2980 (Affirmative) < succ>
 50 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

55 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 941

A DNA sequence (GBSx0999) was identified in *S.agalactiae* <SEQ ID 2861> which encodes the amino acid sequence <SEQ ID 2862>. Analysis of this protein sequence reveals the following:

```

Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3548 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 942

A DNA sequence (GBSx1000) was identified in *S.agalactiae* <SEQ ID 2863> which encodes the amino acid sequence <SEQ ID 2864>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1685 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9525> which encodes amino acid sequence <SEQ ID 9526> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2865> which encodes the amino acid sequence <SEQ ID 2866>. Analysis of this protein sequence reveals the following:

```

Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3125 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 179/476 (37%), Positives = 279/476 (58%), Gaps = 5/476 (1%)

```

Query: 1 MRIEALMGKERRVQYRLLSFLRGSFQALAKLALLTGLSRATFLKYINMLNSYFQSKV 60
M+IE IM+KERR QYRLL L + + + LK + + L$+ T LKTI+NLN ++
Sbjct: 21 MKIEDLMOKERRAQYRLLVFLYHAKNTLRKDLNRLSNLSKVLKLYIDMLNHLCREQGL 80

Query: 61 MCRIVVYKDKLFLFEDYNLSNQEVLLKALMKDSIKYITLILSFNCRQFTIVGLSCLAVSE 120
C++ RD L L+E+ +++ L+K+S+ Y IL ++ F I LS ELAVSE
Sbjct: 81 ACQLLEKDSLSLKNGQPHWEDVALLKESVAYQLTYMYCHEPNTINLSVELAVSE 140

Query: 121 ATLNRHLAHLAEELLAEFDIALSQGKIGDELQWRYFFYLFPQLMSYDKCQNMKKKLDLD 180
ATLNR LAHLN+LL+EFD+A+SQG+Q+G ELQWRYFY+ELP+ + ++ +LD

```

-1033-

Sbjct: 141 ATINRQIAHLNQLSEFDLALSCGSHQLGSRILQWRYFYFPLFRHTLTQSGIDALVNQLDAS 200

Query: 181 SLILATERLAQNTLTREAHQNLGLWFSICHRLAMEKISDNLKPVVKHYQCNAPYKRLD 240
 L LIERL +L+ EA + L +W +I R+ + +D+ N F+KRL+

5 Sbjct: 201 HLA TLIERLIGQSLSAEALSQLTLWLAISQARMSFQKSYNDHFLRSDPMTSNIFFPKRLS 260

Query: 241 AALVLYMSRFALKYREGKVLATFAPLHSCNTLPINTIMEYMGFGGPIIDCVETIIFYFKK 300
 + L+ Y+ R+ALE+ E + F FLH+ +LPI +N+Y +GGGPI D ++E + KK

10 Sbjct: 261 SMLLIYLRRYALEFPAFSAKSLFVFLHAYPLPLPIASKYSLGGPGGPIADHISEALMLKK 320

Query: 301 ESLADETSQVIYQLGQLYSHYFFKGHILVQPDLEQTVRLIDHNRDKLHHSIKCI 360
 ++ +T +++IY LG +S YFFKG IL + + Y+L+ + R L I ++

15 Sbjct: 321 AHVITHQTKERIITYGLGIFPFSKAYFPKSAILSQPTNSQYLYQLVGEDKRALLRVINHLV 380

Query: 361 ANVNRIRPLTEDGCSLLTLHLLELLIFSKNSQKMPFRIGLDMTCNAVEQSLLVIRGHF 420
 +++ D L+ +L LLIPS P +GL + N VE ++ E IR+H

20 Sbjct: 381 LQMDQ----KTDFSQQLSDUILLILFSTKRHHEPLLVLGALGQNKVEAAIAHLAIRRH 436

Query: 421 SGNSISIQVEPYDEGKGFDMVYQSGHSRYPKAKLYCLMKGASERLEQESLLDY 475
 Q+ PYD K +D ++ YQ+ P + Y L + +S EL +++ + D

Sbjct: 437 GHRDFQLMPYDHQKVYDCLITYQTVCLPRQDLFYRLAQYSSPYRLTALEAFKND 492

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 943

A DNA sequence (GBSx1001) was identified in *S. agalactiae* <SEQ ID 2867> which encodes the amino acid sequence <SEQ ID 2868>. This protein is predicted to be transketolase (tktA-I). Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.2084 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9527> which encodes amino acid sequence <SEQ ID 9528> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:BA06071 GB:AP001515 transketolase [Bacillus halodurans]
 Identities = 403/661 (60%), Positives = 520/661 (77%), Gaps = 8/661 (1%)

Query: 6 IDQLAVNTVRLTSLIDAIQAANSCHPQLPMGAAPMAYVLANKPLAVNPKTSRINWTNRDRFV 65
 ++QLAVNT+RTLSID+++ ANSGHPG+PMGAAPMA+ LM KP+N NP + +W NNRDRFV

45 Sbjct: 5 VEQLAVNTVRLTSLIDSEVKNSCHPQPMGAAPMAFCLATKPMNHPN-ANPDWNRDRFV 63

Query: 66 LSAGHGSAILLYSLHLIAGYDLSIDDLKQPRQWSKTPGHPSVNHITGVEATTPLOQGGIA 125
 LSAGHGS LLYSLHLI GYDLS+++L+ PRQWSKTPGHPE HT GVEATTPLOQGA++LL

50 Sbjct: 64 LSAGHGSMLLYSLHLITGYDLSLEELQNPQWSKTPGHPEYGHITGVEATTPLOQGA 123

Query: 126 NAVGMAMARHAHLAAKFNKPGFDLVDRHYTTTLHGDCGLMGVSGEASLAGHLKGLKLVLL 185
 AVGMAMAR HLA A +N G+++VDHYTTT+ GDG LMGVSG EASLAGHLKGLG++LL

55 Sbjct: 124 NAVGMAMARHAATYNRDQYNIVDHYTTTTCGLDGLMGVSGARASLAGHLKGLSMILL 183

Query: 186 YDSNDISLDGPTSQSFTEDVKGRFESYGNQHILVKDGNLEATAAIEAAKAEKTKPTII 245
 YDSNDISLDG SF+E V+ RF++YGN + V+DGN+L+ IA AIE AKA++ +F++I

Sbjct: 184 YDSNDISLDGLHSPSSVDREFKAYGNHVRVEDGNLDELAKAIEAKAD-EPSPIL 242

Query: 246 EVKTIIGPKAEKQGTSSV-HGAPLGARITFAKKAYVWEYP-DFTVEAEVADRFASDLQA 303

-1034-

```

      EVKT IGF+ +G SV HGAPLGA+ + K+AY W Y +F +P EVA + ++
Sbjct: 243 EVKTTIGRSGENKGGKSVSHGAPLGADEVKLTKEAYEWYENEPHIEEVA-AYTEQVKQ 301

Query: 304 RGAKEAEANWDLFAKYEVPELAYEYKEAFAG---QARTVELKAHDLGSSVSRVSSQQ 360
      +GA+ KE+WN+LFA+Y+ YPELA++++ A G + +++G SVA+R SS +
5 Sbjct: 302 QGAKEEESWELFAQYKQAYPELASQPELAVHGDLPKGDVAVPSYEVGKSVATRSSSGE 361

Query: 361 AIQQLSTQLPNLWGGSDLSASNNWVAETDFOASNYAGHNWFGVREFANAAAMNGIA 420
      A+ + +P L+GGSADL+SN T++ E +F +Y+GRN+WFGVREFAN AAMNG+A
10 Sbjct: 362 ALNAFAKTVPLQFGGSADLASSNKLTKGEANFSRDDYSGRNWFGVREFAMGAAMNGA 421

Query: 421 LHGGTRVYGGTFPVFSNYLLPAVRMAALQNLPVYVMTHDSIAVGEDGPTHEPIQLASV 480
      LEHG +V+G TTFVFS+YL FA+R+AAL LE +YV THDSIAVGEDGPTHEP+EQLAS+
15 Sbjct: 422 LHGGLKSVFGATFFVFSYLRPAIRLAALMLPVIYVTHDSIAVGEDGPTHEPVEQLASL 481

Query: 481 RSMPLNLRVIRPADGNETHAAMQRAVSEIDRPTMLVLRQNLPVLEGTSELAQGVNKGAY 540
      R+MP L+VIREADGNE+ AAM+ A+ D+PT LVL+RQNLP LEG +A +GV+KRAY
Sbjct: 482 RAMPGLSVIRPADGNESVAAWKLALESQDPTALVLSRQNLPLEGAVDRAVDGVS KRAY 541

Query: 541 ILSEAKGELDGLIIATGSEVKALDITQDKLESEGIHVRVVSMPACNIFDQEAQEYQVQL 600
      +L+ A G D +++A+GSEV LA++ ++ LE EGIH VVSMP+ + F+ Q A Y+E+VL
20 Sbjct: 542 VLAANGSADLLLLASGSEVLAVNAKEALEKEGIHRAVVSMPDWRFEQAAYKEEVL 601

Query: 601 PSAVTKRLATEAGSGSPWGKYVGLNGLTLTIDTWGASAPGNRI FEEYGTVENAVSLYKEL 661
      PS VT RLATE GSS GW KYVG G + ID +GASAPG RI EE+GPTV++ V+ K L
25 Sbjct: 602 PSDVTARLATEMGSSLOWAKYVGNQGDVVAIDRFGASAPGERIMEEFGPTVQHVVARAGL 662

```

There is also homology to SEQ ID 520.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 944

A DNA sequence (GBSx1002) was identified in *S. agalactiae* <SEQ ID 2869> which encodes the amino acid sequence <SEQ ID 2870>. Analysis of this protein sequence reveals the following:

```

Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4477 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
40      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9529> which encodes amino acid sequence <SEQ ID 9530> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2871> which encodes the amino acid sequence <SEQ ID 2872>. Analysis of this protein sequence reveals the following:

```

Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4581 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
50      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 27/79 (34%), Positives = 45/79 (56%)

-1035-

Query: 3 MKKCKRDFYRQIQHTYNDISVREDAVLSSILLASNGLIKTSOVPRVAYELTQOLENNEI 62
 M+K+ + Y I+ Y+ RE+ LS +LL+ASN LK S+ VAY+L Q +N+ +
 Sbjct: 1 MEKKRQLYDVIRQAYDYPIENRENVALSQLLLAASNRLIKHSNPLLVAYQLWQDVDRYLL 60

Query: 63 EKSFESLATVKELEKSAK 61
 + ++ K+S +K
 Sbjct: 61 DNDILLPKSLCRPKQSLK 79

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 945

A DNA sequence (GBSx1003) was identified in *S. agalactiae* <SEQ ID 2873> which encodes the amino acid sequence <SEQ ID 2874>. This protein is predicted to be ABC transporter, ATP-binding protein.

- 15 Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 20 bacterial cytoplasm --- Certainty=0.2610 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 25 >GP:CA849925 GB:AJ248286 ABC transporter, ATP-binding protein
 [Pyrococcus abyssi]
 Identities = 96/243 (39%), Positives = 164/243 (66%), Gaps = 2/243 (0%)

- Query: 1 MIKFEHVSKVYGEKEALSDLTLSVKDGEIFGLIGHNGAGKTTTISILTSIIDATYGVYI 60
 MI E++ K +G KE L ++ +VKDGEI+GL+G NG+GK+TT+ IL+ II G+V +
 Sbjct: 1 MIIVENLRKRFPGGKEVLKGISFTVKDGEIYGLLGPNGSGKSTTMRILSGIITDFEGKVIV 60

- Query: 61 DDILLTEHRDQIKKIKGIVPDSDFINLTAREYMYFLAKIYDVAPEDIEARITKLVDIF 120
 + + + Q+K+ +GYVP+P ++ +LT E++ F+ + + +E R+ KLV+ F
 Sbjct: 61 GGVEVAKDPLQVGRIVGYVPETPALYBSLTPAEFFSFVGGVGRGIPKDIIEERVRLKLEAF 120

- Query: 121 ELEEQRYPNIESFSHGMROKIVIVIGALLNFDIWLDEPLTGLDQASFDLKEMKEHAK 180
 E+++ I + S G +QK+ +I +LL +P + ILDE + GLDP+++ +E++ E +
 Sbjct: 121 EIKYNNQLITGLTSLFGTKKISLISLHDPKVLILDEAMNGLDPSARIFRELLVEPK 180

- Query: 181 NGKTVIFSTHVLAVABQLCDRIGILKQKLI FVGSIGELKNKYDPKDLITYILELAGRQA 240
 GK+++FSTHVLAAE +CDR+GI+ QG+I G++ ELK ++ LE ++L+L QA
 Sbjct: 181 EGKSIVSTHVLAAELICDRVGIIYQGRIGABGTVEELKEISGHERLEDVFLKLT--QA 238

- 45 Query: 241 SRE 243
 S
 Sbjct: 239 KEE 241

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2875> which encodes the amino acid sequence <SEQ ID 2876>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 55 bacterial cytoplasm --- Certainty=0.2723 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-1036-

Identities = 182/244 (74%), Positives = 215/244 (87%)

Query: 1 MIKPEHVSXVYGSKEALSOLTSVKGDEIPGLIGHNGAGKTTTISILTSIIDATYGVYI 60
 NI+P+HVSX+Y+G+KEALSOL+ ++ DGEIPGLIGHNGAGKTTTISII+Y+V+V+
 5 Sbjct: 1 NIEFHVSKLYGDKALSDLVNTINDGEIPGLIGHNGAGKTTTISILTSIEASYGEVVFV 60

Query: 61 DDLLETHEDQIKKIGVVPDSDFILNLTAREYWPFLAKIYDVAPEDIEARITKLVDIF 120
 D LITE+R+ IKK+I YVDSDFILNLT EYN FLAKIY V+ ED E R+ +L +F
 10 Sbjct: 61 DQQLTENRZAIIKKIAYVPDSDFILNLTAREYWPFLAKIYGVSDERERRLAQUTTLF 120

Query: 121 ELERQRYNPDESFSHGMRQKVIVIGALLPNPDIWLIDEPFLGLDPQASFDLKEEMKHAH 180
 EL+E+ I+FSHGMRQKVIVIGAL+ NP+IWLIDEPFLGLDPQASFDLKEEMKHA
 10 Sbjct: 121 ELKREVNQYIDSFSHGMRQKVIVIGALVSNPNIWLIDEPFLGLDPQASFDLKEEMKHAHA 180

Query: 181 NGRTVIFSTHVLAVARQLCDRIGILKQKGLIFVGSGLGELRMKYPDKDLETIYLELAGRQA 240
 +G TV+FSHVL+VARQLCDRIGILK+GKLI+PVG++ ELK +PDKDL+IYLELAGR+R+
 15 Sbjct: 181 SGRTVLFSHVLVARQLCDRIGILKQKGLIFVGTIDELKEHPDKDL+IYLELAGRKA 240

Query: 241 SREG 244
 EG
 20 Sbjct: 241 QREG 244

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 946

A DNA sequence (GBSx1004) was identified in *S.agalactiae* <SEQ ID 2877> which encodes the amino acid sequence <SEQ ID 2878>. Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence

30	INTEGRAL	Likelihood = -13.43	Transmembrane	504 - 520 (495 - 529)
	INTEGRAL	Likelihood = -12.58	Transmembrane	427 - 443 (400 - 449)
	INTEGRAL	Likelihood = -10.99	Transmembrane	151 - 167 (144 - 179)
	INTEGRAL	Likelihood = -8.44	Transmembrane	194 - 210 (189 - 214)
	INTEGRAL	Likelihood = -7.96	Transmembrane	48 - 64 (46 - 68)
35	INTEGRAL	Likelihood = -7.32	Transmembrane	350 - 366 (348 - 378)
	INTEGRAL	Likelihood = -6.69	Transmembrane	475 - 491 (474 - 501)
	INTEGRAL	Likelihood = -6.00	Transmembrane	319 - 335 (318 - 337)
	INTEGRAL	Likelihood = -5.73	Transmembrane	252 - 268 (244 - 271)
	INTEGRAL	Likelihood = -4.78	Transmembrane	125 - 141 (121 - 148)
40	INTEGRAL	Likelihood = -4.51	Transmembrane	76 - 92 (71 - 98)
	INTEGRAL	Likelihood = -3.56	Transmembrane	406 - 422 (400 - 426)

----- Final Results -----

45	bacterial membrane	--- Certainty=0.6371 (Affirmative) < succ>
	bacterial outside	--- Certainty=0.0000 (Not Clear) < succ>
	bacterial cytoplasm	--- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2879> which encodes the amino acid sequence <SEQ ID 2880>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

55	INTEGRAL	Likelihood = -14.33	Transmembrane	167 - 183 (158 - 193)
	INTEGRAL	Likelihood = -12.52	Transmembrane	524 - 540 (508 - 546)
	INTEGRAL	Likelihood = -10.93	Transmembrane	63 - 79 (60 - 84)
	INTEGRAL	Likelihood = -8.39	Transmembrane	421 - 437 (414 - 456)
	INTEGRAL	Likelihood = -8.23	Transmembrane	208 - 224 (203 - 228)
	INTEGRAL	Likelihood = -8.23	Transmembrane	504 - 520 (493 - 521)
	INTEGRAL	Likelihood = -7.59	Transmembrane	139 - 155 (134 - 162)
60	INTEGRAL	Likelihood = -6.64	Transmembrane	261 - 277 (257 - 287)

-1037-

INTEGRAL	Likelihood = -4.99	Transmembrane	446 - 462 (444 - 464)
INTEGRAL	Likelihood = -4.25	Transmembrane	369 - 385 (367 - 387)
INTEGRAL	Likelihood = -0.80	Transmembrane	87 - 103 (87 - 104)
INTEGRAL	Likelihood = -0.11	Transmembrane	334 - 350 (334 - 350)

----- Final Results -----

bacterial membrane --- Certainty=0.6731(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9173> which encodes the amino acid sequence <SEQ ID 9174>. Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -14.33	Transmembrane	153 - 169 (144 - 179)
INTEGRAL	Likelihood = -12.52	Transmembrane	510 - 526 (494 - 532)
INTEGRAL	Likelihood = -10.93	Transmembrane	49 - 65 (46 - 70)
INTEGRAL	Likelihood = -8.39	Transmembrane	407 - 423 (400 - 442)
INTEGRAL	Likelihood = -8.23	Transmembrane	194 - 210 (189 - 214)
INTEGRAL	Likelihood = -8.23	Transmembrane	490 - 506 (479 - 507)
INTEGRAL	Likelihood = -7.59	Transmembrane	125 - 141 (120 - 148)
INTEGRAL	Likelihood = -6.64	Transmembrane	247 - 263 (243 - 273)
INTEGRAL	Likelihood = -4.99	Transmembrane	432 - 448 (430 - 450)
INTEGRAL	Likelihood = -4.25	Transmembrane	355 - 371 (353 - 373)
INTEGRAL	Likelihood = -0.80	Transmembrane	73 - 89 (73 - 90)
INTEGRAL	Likelihood = -0.11	Transmembrane	320 - 336 (320 - 336)

----- Final Results -----

bacterial membrane --- Certainty=0.673(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 255/542 (47%), Positives = 378/542 (69%), Gaps = 12/542 (2%)

Query: 1 MWNSIWELKINILYSNPQTLSALRKKQEKHPKGFSAKMSFNRQLPQLLPSLIYVF 60
 MWNS IWELKINILYSNPQ+L+ L+K+QEKHPK+ F AYKSM R Q I +F +Y+V+
 Sbjct: 15 MWNSTIWELIKINILYSNPQSLANLKKRQEKHPKCNFKAYKSMRQALMIAMFLVLIYLF 74

Query: 61 LFSVLDPKFEPYGPYPTFYIGIFLVSIIYSFIAMYSVFPYSDDVQYAYLPIKSEELYAK 120
 +F+ +DF YPG P+P+ +F +S+ +F ++Y++FYES+D+K Y +LP+ SEELY+AK
 Sbjct: 75 MFIGVDFSHYPGLPSFDVAMFPFIMSTLTAFSSLYTIFYESNDLKIYHLFVTSSEELYAK 134

Query: 121 IFATPGMSVTFMLPILTLMIYVAYNRHIGGPLAVLLAIINFALLFSLVTVISLYINSLIGK 180
 I+ +GM FLMP++L++AYN++G PL++L+AI+ F+L+S V++YIN+ +G+
 Sbjct: 135 IVSSLGNGAVFLMPLISLLLLAYWQLGNPLSLIVLIVFLVLSVSSMLAIYINAWGK 194

Query: 181 AIIRSANRKLITSLISLAFPGAIVLFLFVWTSQK--WVQKQLQDIAPFYVGVYDIV 238
 I+RS RKLITI++ ++TFGA V+ +N+++ K M G D IPY +G+YD+V
 Sbjct: 195 IIVRSRRKLITIMMPVSTFGAVFLIYAINISNNKRIWTDGVFTDYPTIPYKPGYDVV 254

Query: 239 TAPFSMESLNYLLELILFLIGALYKWMERYQSELY---GQVKORK--VHQKIDF 292
 APPS +L+N++LPL+IL++ I WVP YY+E Y +VQK R +R
 Sbjct: 255 QAPPSTALLNFWLPLLLILAMVYIVTKWMPYTYREAPYISNNKRVQTKKPVNR--- 311

Query: 293 SKRESINKTLVXHHLSLQNA+LITNTFIMPLLYLAMFIVDILNNGKEIGRFNNFYGI 352
 +S+ +L KHL +LQNA+LIT T+LMP+Y+ +FI P L+ G + + +YPG+
 Sbjct: 312 HQNQLAQLLRKHLLTLQNA+LITQTYTLMPLMYMLFVIGPSLSRGTGPFKHISPDYPGV 371

Query: 353 AFLAGLIGSLCVMFASIVGVGISLEKSNFYIKSLPIGSPYFLKHGFPVTLITLQLAVPT 412
 A L G+ +G+ C P S +GVGISLEK NF FIKSLPI+ FL KF L+ LQL VP
 Sbjct: 372 ALLFGVSLGVNCAFTPTSFVGVISLEKINFTYIKSLPTLAKFLMDKFCLLVGLQLIVPM 431

Query: 413 PIYFLVAGFLLKSLILVLLSFLGLWFLKLTGQPIYRRDYKHLPLAMQVETQLFNRLGL 472
 IY + G F+L L L+ ++P LG +++G+ +YRRDY+ L L WQ++TQLF RG G

-1038-

Sbjct: 432 VIYLVPGLFVLHLHPLLTIAFCGLYALSLVQGHMYRDRYRLDLKQDWTQLPTRGD 491
 Query: 473 QWLLAKSLFGNMIIIGSFLL-IGISIFWSSWVWVAVNIIILIGLLISIQYLLALNFWK 531
 QWL +G +FG +I+ L G I +++ + ++I+ + L+L+ Q + K FWK
 Sbjct: 492 QWLLWGLIFGNLIVAGVLFSGAVIAMIQQPLLISILLISLTIAMVLSLAQLWIKTFWK 551
 Query: 532 KL 533
 L
 Sbjct: 552 SL 553

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 947

A DNA sequence (GBSx1005) was identified in *S. agalactiae* <SEQ ID 2881> which encodes the amino acid sequence <SEQ ID 2882>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.12 Transmembrane 242 - 258 (239 - 265)
 INTEGRAL Likelihood = -7.64 Transmembrane 430 - 446 (421 - 450)
 INTEGRAL Likelihood = -5.84 Transmembrane 120 - 136 (113 - 139)
 INTEGRAL Likelihood = -5.52 Transmembrane 212 - 228 (210 - 232)
 INTEGRAL Likelihood = -5.20 Transmembrane 287 - 303 (283 - 313)
 INTEGRAL Likelihood = -3.56 Transmembrane 148 - 164 (143 - 166)
 INTEGRAL Likelihood = -0.48 Transmembrane 382 - 398 (382 - 398)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15963 GB:Z99124 phosphotransferase system (PTS)
 beta-glucoside-specific enzyme IIABC component [Bacillus subtilis]
 Identities = 175/447 (39%), Positives = 266/447 (59%), Gaps = 10/447 (2%)
 Query: 4 EYITLSKNIKKLGGQNMNNVYHCQTLRPSLNDPTKWLSQLTKLEKTVVI89QGH 63
 +Y LSK+I++ +GG+ N+ V HC TRLRF+L+D K + QI+ L V I89 Q
 Sbjct: 2 IDYKLSKDIQLVGGSENVQVRVHCNTRLRFLNHLNKAADKASQLBQLGPFVMTNIG8QGF 61
 Query: 64 QIVGITHVAKVPEEL---NSLIETNSTTKIQCTKAKAVSRIDFVSCTFQPIPLPLOGA 120
 QI+IG V KV++ I ++L+ S Q K +8 + D +8G F PILPA++GA
 Sbjct: 62 QIIIGNDPVKVYQATVRHSLNLSDEKSAGSSSQKQV--LSAVFDVIGSVFTPIPLAAGA 119
 Query: 121 GMKALLALLVFKILTPSSQTYILNLFADGVFPYLPILILAITAQLKANPILAGTV 180
 GMK L+AL + F + SQ ++L DG FYFLP+L+R++R+K +NP +A
 Sbjct: 120 GMKGLVALAVTFPGWAEKSGVHVILTAVDGCAPYFLPLLAMSAARKFGSNPVAAIA 179
 Query: 181 VMLLHNWANIIVASGKPVSLPFTPTPTLTYASSVPIILLICQVYLEKYLKOIPIKL 240
 +LHP+ L+ +GKP+8 F +P T Y+S+VIDI+L I + +Y+RK+ + SL
 Sbjct: 180 AAILHFDLTALLGKFPIS-FIGLPVTAATYSSTVPIILLIWIYASVEKWDIRFTHSL 238
 Query: 241 RMLVWPMILFLSMGILSPSLGPMCTIAQKLAVIFTFLEKYSASW-APAPLVGAPAPILI 299
 +L+VP L + L+ +GP+G I G+YL+ +L +A A FL G F+ ++I
 Sbjct: 239 KLIVVPTFTLLIVVPLTLITVGPLGAILGEYLSGVMYTFDHAGLVAMIFLAGTF8-LIT 297
 Query: 300 MFCVH8SIAALGITQAKLVDSIIFQPMCLSNIAQNTAGTVVTLITKEKKLKRITAGPRA 359
 M G+H + I +A+ G D + P M +N+ QA A V L + KK K +A +
 Sbjct: 298 MTGMEYAFVPIININIAQNHDIYLL-PAMFLANNQAGSFAVFLSRNKKFKSLALITS 356
 Query: 360 ITAPMGITTEPLYGVNLEKRYPLIASLGGGLGGLYAGIDNAHREAV-GSSGLGILFLYT 418
 ITA MGITTEP +YGVN+ + P A+LIGG GG + G+ + V G+GLP ++I
 Sbjct: 357 ITALMGITTEPAMYGVNMLRKLFPAAALIGGAGGAFGYMGVASTYVGGNGLPSIPVFI 416

-1039-

Query: 419 SRTSTHLFITMLIAVITVSTTALTTF 445
 T + I ++IA S +L F
 Sbjct: 417 GPTFYAMIGLVIAFAAGTSAAYLLGF 443

There is also homology to SEQ ID 2884.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 948

A DNA sequence (GBSx1006) was identified in *S.agalactiae* <SEQ ID 2885> which encodes the amino acid sequence <SEQ ID 2886>. This protein is predicted to be gamma-glutamyl kinase (proB). Analysis of this protein sequence reveals the following:

Possible site: 58
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.11 Transmembrane 160 - 176 (160 - 176)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1044 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA63147 GB:X92418 gamma-glutamyl kinase [Streptococcus thermophilus]
 Identities = 200/265 (75%), Positives = 235/265 (88%)
 Query: 1 MKRHFETTRRIVKVGTSSELVOTSGKINLSKIDHLAFVSSIANRGMVIELVSSGAMGPG 60
 MKR+R++ +R+VIK+GTSSLV SGKINL KID LAFVSSL N+G+EV+LVSSGAMGPG
 Sbjct: 1 MKRNFDSVKRLVIKIGTSSLVSGKINLEKIDQLAFVSSLHNKIEVIELVSSGAMGPG 60
 Query: 61 LDILMKDKRPOEISQQQAVSSVGQVAMMSLYSQIPSHYQTHVSQLLTRDVVVFPESLQN 120
 L++L ++KRP E+ +QQAVSSVGQVAMMSLYSQ+PSHYQT VSQ+LLTRDVV + ESL N
 Sbjct: 61 LNVLLDKRPAEIVGKQAVSSVGQVAMMSLYSQVPSHYQTKVSQLLTRDVVEYSLEIAN 120
 Query: 121 VTNSFESLLSMGILPIVMENDAVSVDEMHDHTKFGENDRLSAVVAKINADLLIMLSID 180
 N+FSLS +G++PIVMENDAVSVDEMHD TKFGENDRLSA+VAK+ ADLLIMLSID
 Sbjct: 121 ATRAFESLFLGVVPIVMENDAVSVDEMHDATKFGENDRLSAIVKVVAGDLLIMLSID 180
 Query: 181 GLFDKNPNYIDAVLRSHVSEITDDIIKSAGGAGSKFPTGMLSKIKSAQNVFNDNGQM 240
 GLFDKNPN+Y+DA LRS+V EIT+I+ SAGGAGSKFPTGGM+SKIKSAQNVF+N QM+
 Sbjct: 181 GLFDKNPNVEDATLRSYVPTETTERILASAGGAGSKFPTGGMMSKIKSAQNVFNSQMV 240
 Query: 241 LMNGENPRDILKVLGDHMGISTYFAQ 265
 LMNG NPRDIL+VL+G IGT F Q
 Sbjct: 241 LMNGENPRDILRVLGKAGITAFKQ 265

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2887> which encodes the amino acid sequence <SEQ ID 2888>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.97 Transmembrane 163 - 179 (163 - 179)
 INTEGRAL Likelihood = -0.06 Transmembrane 124 - 140 (124 - 140)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1786 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

-1040-

>GP:CAA63147 GB:X92418 gamma-glutamyl kinase [Streptococcus thermophilus]
 Identities = 212/265 (80%), Positives = 237/265 (89%)

Query: 4 MKRQFEDVTRIVIKIGTSSLVPTGKINLEKIDQLAFVISSLMNKGKIVLVSSGAMGFG 63
 5 MKR F+ V R+VIKIGTSSLVLP+GKINLEKIDQLAFVISSL NKG EV+LVSSGAMGFG
 Sbjct: 1 MKRNFDVVKRLVIRIKIGTSSLVLP+GKINLEKIDQLAFVISSLMNKGKIVLVSSGAMGFG 60

Query: 64 LDILKMKRPTNLAKQAVSSVGQVAMMSLYSQIFAYQYQTNVSQLLTRDVVVPFESLN 123
 L++L +EKRP + KQAVSSVGQVAMMSLYSQIF++YQT VSQLLTRDVV + ESLN
 10 Sbjct: 61 LNVLDLEKRPABVVGKQAVSSVGQVAMMSLYSQIFSHYQTKVSQLLTRDVVYESLN 120

Query: 124 VTNAFESLISLGIVPIVNENDAVSVDMDHATKFGDNRLSAVAVGATKADLLIMLSOID 183
 NAFESL LG+VPIVNENDAVSVDMDHATKFGDNRLSA+VA + ADLLIMLSOID
 Sbjct: 121 AINAFESLISLGIVPIVNENDAVSVDMDHATKFGDNRLSAITAVKVGADLLIMLSOID 180

Query: 184 GLFDKNPTIYEDAQLRSHVANITQEIISAGGAGSKPGTGGMLSKVQSAQMVPFNKQMW 243
 GLFDKNP +YEDA LRS+V IT+EI+ASAGGAGSKPGTGGM+SK++SAQMVPFN+ QMW
 Sbjct: 181 GLFDKNPNVYEDATLRSYVPIITEIISAGGAGSKPGTGGMMSKIKSAQMVPFNQSQMW 240

Query: 244 LMNGANPRDILRVLEGQPLGTWFFQ 268
 LMNG NPRDILRVLEG +GT FFQ
 20 Sbjct: 241 LMNGANPRDILRVLEGAKIGTLFFQ 265

An alignment of the GAS and GBS proteins is shown below.

25 Identities = 217/265 (81%), Positives = 242/265 (90%)

Query: 1 MKRHFETTRIVIKIGTSSLVPTGKINLEKIDQLAFVISSLMNKGKIVLVSSGAMGFG 60
 MKR FE RIVIK+GTSSLV +GKINL KID LAFVISSLMN+G EVILVSSGAMGFG
 30 Sbjct: 4 MKRQFEDVTRIVIKIGTSSLVPTGKINLEKIDQLAFVISSLMNKGKIVLVSSGAMGFG 63

Query: 61 LDILKMKRPTNLAKQAVSSVGQVAMMSLYSQIFAYQYQTNVSQLLTRDVVVPFESLN 120
 LDILKMK+KRP +++QAVSSVGQVAMMSLYSQIF++YQT+VSQLLTRDVVVPFESL N
 Sbjct: 64 LDILKMKRPTNLAKQAVSSVGQVAMMSLYSQIFAYQYQTNVSQLLTRDVVVPFESLN 123

Query: 121 VTNSFESLISLGIVPIVNENDAVSVDMDHATKFGDNRLSAVAVGATKADLLIMLSOID 180
 VTN+FESL+S+GI+PIVNENDAVSVDMDH TKGDNRLSAVVA ITKADLLIMLSOID
 35 Sbjct: 124 VTNAFESLISLGIVPIVNENDAVSVDMDHATKFGDNRLSAVAVGATKADLLIMLSOID 183

Query: 181 GLFDKNPTIYEDAQLRSHVANITQEIISAGGAGSKPGTGGMLSKVQSAQMVPFNKQMW 240
 GLFDKNP IT+DA LRSHV+ IT +II SAGGAGSKPGTGGMLSK++SAQMVP+N QGM+
 40 Sbjct: 184 GLFDKNPTIYEDAQLRSHVANITQEIISAGGAGSKPGTGGMLSKVQSAQMVPFNKQMW 243

Query: 241 LMNGANPRDILRVLEGQPLGTWFFQ 265
 LMNGANPRDIL+VL+G +GT+F Q
 45 Sbjct: 244 LMNGANPRDILRVLEGQPLGTWFFQ 268

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 949

50 A DNA sequence (GBSx1007) was identified in *S. agalactiae* <SEQ ID 2889> which encodes the amino acid sequence <SEQ ID 2890>. This protein is predicted to be unnammed protein product (proA). Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3517 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

60

-1041-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2891> which encodes the amino acid sequence <SEQ ID 2892>. Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA63148 GB:X92418 gamma-glutamyl phosphate reductase
 [Streptococcus thermophilus]
 Identities = 309/416 (74%), Positives = 355/416 (85%)
 Query: 1 MTDMRRIGORAKQASLLIAPLSTQIKRFLSTLAKALVDDTQTLAAQKDLNAKEHGI 60
 MT + LGQAKAS IA LST KN L+ +AKALV ++ + N KD+ANA E+GI
 Sbjct: 1 MTYVDTLGGQAKVASRQIAKLSTAQNDDLNVAKALVAESDYIFTEAKDMANASENGI 60
 Query: 61 SDIMMRRLRLTSERIKALAQGVQVQADLADPIGVQVIGYTNLDGLKILQKRVPLGVMI 120
 S IM DRL LT +RI IA+GV+QVADL DPIGOV++GYTNLDGLKI+QKRV+GVMI
 Sbjct: 61 SKIMQDRLLLTEDRIAGIABGVQVQADLQDPIGVQVIGYTNLDGLKIVQKRVPMGVMI 120
 Query: 121 FESRPNVSDAFSLAFKTNNAIILRGKKDALHSNKALVKLRQSLKSGITPDVAQVLED 180
 FESRPNVSDAFSLAFKTNNAIILRGKKDALHSNKALV+R++L+ +GIT DAVQ VED
 Sbjct: 121 FESRPNVSDAFSLAFKTNNAIILRGKKDALHSNKALV+R++L+ +GIT DAVQ VED 180
 Query: 181 PSHVAEELMQATDYVDVLI PRGGAKLIQTVKEKAKVPVLETGVGNHVIYDQAQDLDA 240
 SH VAEELM A+ YVD+LIPRGA+LIQTVKEKAKVPVLETGVGNHVIYD A+LD+A
 Sbjct: 181 TSHVAEELMQATKYVDLLI PRGGAKLIQTVKEKAKVPVLETGVGNCHITVQKALNDMA 240
 Query: 241 TKIVINAKTRPSVCNAAEGLVHRAVAARFIPMEKAINQCVQVWRADDKALFLPEQA 300
 T+IVINAK+RPSVCNAAE LV+H + F+P LEKAI++Q VE+RAD++AL L E+A
 Sbjct: 241 TQIVINAKTRPSVCNAAEGLVHRAVADVEEFLPNLEKAIKIQVSEVFEADERALKMEKA 300
 Query: 301 VPAKARDFETEFLOYIMSVKVVSLSEALSWINQYTSHSEALITRDIKAAETPDQLVDA 360
 VPA EDF TEFLDYIMSVKVV SL+RAI+WIN YT+ HSEAL+T+DI AE FQQ VDA
 Sbjct: 301 VPASPEFATEFLDYIMSVKVVSLSEALSWINQYTSHSEALITQDISRAEQFQDDVDA 360
 Query: 361 AAVYVNASTRFTDGFVFLGAEIGISTQKMHARGPMGLEALTSTKPYINGDGHIRE 416
 AAVYVNASTRFTDGFVFLGAEIGISTQKMHARGPMGLEALTSTKPYINGDGHIRE
 Sbjct: 361 AAVYVNASTRFTDGFVFLGAEIGISTQKMHARGPMGLEALTSTKPYINGDGHIRE 416

An alignment of the GAS and GBS proteins is shown below.

Identities = 307/417 (73%), Positives = 353/417 (84%), Gaps = 1/417 (0%)
 Query: 1 MTYIIRILGQAKKASQSVARLSTASKNRIILDLARNIVADTETILTEARNADVVKADNGI 60
 MT + LGQ AK+AS +A LST KN L LA++ V D+T+L N +D+ AK+GI
 Sbjct: 1 MTDMRRIGORAKQASLLIAPLSTQIKRFLSTLAKALVDDTQTLAAQKDLNAKEHGI 60
 Query: 61 SEIMVDRRLRLNKDRIQAIANGIYQVADLADPIGVQVSGYTNLDGLKILKRVPLGVMI 120
 S IM+DRRL +RI+AI+GA QVADLADPIGV+ GYTNLDGLKIL+KRVPLGVMI
 Sbjct: 61 SDIMMRRLRLTSERIKALAQGVQVQADLADPIGVQVIGYTNLDGLKILQKRVPLGVMI 120
 Query: 121 FESRPNVSDAFSLAFKTNNAIILRGKKDALHSNKALVKLRQSLKSGITPDVAQVLED 180
 FESRPNVSDAFSLAFKTNNAIILRGKKDALHSNKALV+R++L+ +G PD VQVLED
 Sbjct: 121 FESRPNVSDAFSLAFKTNNAIILRGKKDALHSNKALVKLRQSLKSGITPDVAQVLED 180
 Query: 181 TSHVAEELMQATDYVDVLI PRGGAKLIQTVKEKAKVPVLETGVGNHVIYDEFAQLDMA 240
 SH VAEELMQATDYVDVLI PRGGAKLIQTVKEKAKVPVLETGVGNHVIY+D ADLQ+A
 Sbjct: 181 PSHVAEELMQATDYVDVLI PRGGAKLIQTVKEKAKVPVLETGVGNHVIYDQAQDLDA 240
 Query: 241 AKIVINAKTRPSVCNAAEGLVHRAVAARFIPMEKAINQCVQVWRADDKALFLPEQA 300
 KIVINAK+RPSVCNAAEGLV+H+A+A F+ LEK + + Q VE+RAD++AL L E

-1042-

Sbjct: 241 TKIVINAKTKRPSVCHAAEGLVIEHVAARFIPMLEKALNQV-QPVENRADDKALFLFBEQ 299

Query: 301 AVAASSDYATEFLDYIMSVKVVDSFBQALSWINKYSSHHSEAITTNISRAEITPQMDVD 360
AV A D+ TFLDYIMSVKVV S E+ALSWIN+Y+SHHSEAIT I AE PQD+VD

Sbjct: 300 AVPAKAEDEFETFLDYIMSVKVVSSLEFAISWINQYTHHSEAITTRDIKAAETPQQLVD 359

Query: 361 AAAYVNASTRPTDGFVFGLGARIGISTQKLHARGFMGLRALTSTKYINGTGQVRE 417
AAAYVNASTRPTDGFVFGLGARIGISTQK+HARGFMGLRALTSTK+YING G +RE

Sbjct: 360 AAAYVNASTRPTDGFVFGLGARIGISTQKMHARGFMGLRALTSTKYINGDGHIRE 416

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 950

A DNA sequence (GBSx1008) was identified in *S. agalactiae* <SEQ ID 2893> which encodes the amino acid sequence <SEQ ID 2894>. Analysis of this protein sequence reveals the following:

Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1859 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9531> which encodes amino acid sequence <SEQ ID 9532> was also identified.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2895> which encodes the amino acid sequence <SEQ ID 2896>. Analysis of this protein sequence reveals the following:

Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0853 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 259/315 (82%), Positives = 287/315 (90%)

Query: 1 MTNDRHITVLLHETVMDLTIKPDGIYVDATLGGAGHSSYLLSQLGPDGHLYAFDQDQKA 60
MT +FHH+TVLLHETVMDLTIKPDGIYVDATLGG+GHS YLLS+LG +GHLY PDQDQKA

Sbjct: 22 NTCSFHHITVLLHETVMDLTIKPDGIYVDATLGGSGHSAYLLSKLGBEGHLYCFDQDQKA 81

Query: 61 IDNAHRLKKYVDTGQVTFIKDNFRNLSNKLALGVSEINGT CYDLGVSSPQDLDERBGRF 120
IDNA + LK Y+D GQVTFIKDNFR+L + L ALGV EI+GI YDLGVSSPQDLDERBGRF

Sbjct: 82 IDNAQVTLKSYIDKGQVTFIKDNFRHLKARLTALGVDEIDGILYDLGVSSPQDLDERBGRF 141

Query: 121 SYKQDAPLDMRMNRQSLTAYVUNVTYSYHDLVRIFPKYGEDKPSQIARKIEQVRAEKT 180
SYKQDAPLDMRM+R+ L TAY+VUNVTY ++DIV+IFPKYGEDKPSQIARKIEQ RA K

Sbjct: 142 SYKQDAPLDMRMNRQSLTAYVUNVTYFNDLVKIFPKYGEDKPSQIARKIEQARAIEK 201

Query: 181 ISTTTLEAEILKSSKACELKKGHPAKQIFQAIRIEVNDELGADESIQAMDELLALDG 240
I TTTLEAE+IK++K AKELKKGHPAKQIFQAIRIEVNDELGADESTQ AN+LL+DG

Sbjct: 202 IETTTLEAEILKAAKPAELKKGHPAKQIFQAIRIEVNDELGADESIQAMDELLALDG 261

Query: 241 RISVITFHSLEDRITKQLPKCASTVVPKGLFPIPDQLPKMKLVNRRKPIILSPSELEAN 300
RISVITFHSLEDRITKQLPKCASTV+VPKGLP IP+D++PK ELV+RKPPIILS EL AN

Sbjct: 262 RISVITFHSLEDRITKQLPKCASTVDVVPKGLFPIPDMLPKFELVSRKPIILSPSELEITAN 321

Query: 301 NRAHSAKLRVARRIR 315
RAHSAKLRVA++IR
Sbjct: 322 KRAHSAKLRVAKKIR 336

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 951

A DNA sequence (GBSx1009) was identified in *S.agalactiae* <SEQ ID 2897> which encodes the amino acid sequence <SEQ ID 2898>. This protein is predicted to be FtsL. Analysis of this protein sequence
10 reveals the following:

Possible site: 42
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -8.92 Transmembrane 30 - 46 (24 - 49)
15 ----- Final Results -----
bacterial membrane --- Certainty=0.4567 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC95455 GB:AF068903 Y1LD [Streptococcus pneumoniae]
Identities = 44/99 (44%), Positives = 71/99 (71%)

- 25 Query: 5 KRTEAVTQTLQRHIKTFSRIEKAFYGAIVITAINAVGIIYQLQNSLQVQEVQVQLNSKI 64
++ E Q LQ +K FSR+EKAFY +I +T +I+A+ II++Q+ LQV+ ++ ++N+I
Sbjct: 3 EKMEKGTGQILQNLKRFSRVEKAFYFSIAVTTILVAISIIIPMQTKLLQVQNDLTAKNAQT 62

- Query: 65 NDKQTFDNAAKQEVNELSNDRITKIAKDALGTIONDNI 103
+K+TE D+AKQEVNEL +R+ +IA L +N+NI
30 Sbjct: 63 EEKKTLEDAAKQEVNELRAERLKEIANSHDLQLANNENI 101

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2899> which encodes the amino acid sequence <SEQ ID 2900>. Analysis of this protein sequence reveals the following:

Possible site: 50
35 >>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -5.79 Transmembrane 40 - 56 (37 - 58)
----- Final Results -----
bacterial membrane --- Certainty=0.3314 (Affirmative) < succ>
40 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

- >GP:AAC95455 GB:AF068903 Y1LD [Streptococcus pneumoniae]
45 Identities = 45/94 (47%), Positives = 69/94 (72%)
Query: 24 LQGRKINTFSRIEKAFTTAAIVITAINAVSIIYQLSRKQLQEQEITSLNSHISDQKLELN 83
LQ ++K FSR+EKAFY +I VI + +A+SII++Q++ LQ+Q ++T +N+ I ++K EL++
50 Sbjct: 12 LQMQLKRFSRVEKAFYFSIAVTTILVAISIIIPMQTKLLQVQNDLTAKNAQIEEKTLED 71
Query: 84 AKQEVNELSRDRIDIAGKAGLSNRNNHIKKVE 117
AKQEVNEL R +R+ +IA L N NI+ E
Sbjct: 72 AKQEVNELLRABRLKEIANSHDLQLANNENIRIAE 105

- 55 An alignment of the GAS and GBS proteins is shown below.

Identities = 71/108 (65%), Positives = 87/108 (79%), Gaps = 1/108 (0%)

-1044-

Query: 1 MTNKKRTQAVTQTLQRHITKTFSRIEKAFYGAIVITAINAVGIIYLCNSLQVKQBNQL 60
 MTNKKRT+ VT LQ+ IKTFSRIEKAFY AI++TAL MAV IYLCNS LQ++QE+ L
 Sbjct: 11 MTNKKRTQVINALQKRITKTFSRIEKAFYTAIIVTAITNAVSIIYLCNSRLQLQEQITSL 70

5 Query: 61 MSKINDKQTEFDNAKQEVNLSNRDRITTKIAKDAGLTQNDNIYRKVD 108
 MS I+D++ E +NAKQEVNLS RDRI IA AGL+ +N+NI +KV+
 Sbjct: 71 NSHISDQKLELNNAKQEVNLSNRDRITTDIAGKAGLSNRNNI -RKVE 117

SEQ ID 2898 (GBS82) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell
 10 extract is shown in Figure 15 (lane 2; 2 bands).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 952

A DNA sequence (GBSx1010) was identified in *S. agalactiae* <SEQ ID 2901> which encodes the amino
 15 acid sequence <SEQ ID 2902>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 20 bacterial cytoplasm --- Certainty=0.1435(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

25 No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 vaccines or diagnostics.

Example 953

A DNA sequence (GBSx1011) was identified in *S. agalactiae* <SEQ ID 2903> which encodes the amino
 30 acid sequence <SEQ ID 2904>. This protein is predicted to be unnamed protein product. Analysis of this
 protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood =-13.90 Transmembrane 37 - 53 (30 - 60)

----- Final Results -----
 35 bacterial membrane --- Certainty=0.6562(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2905> which encodes the amino acid
 sequence <SEQ ID 2906>. Analysis of this protein sequence reveals the following:

Possible site: 42
 45 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood =-13.06 Transmembrane 33 - 49 (24 - 53)

----- Final Results -----
 50 bacterial membrane --- Certainty=0.6222(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1045-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 480/753 (63%), Positives = 603/753 (79%), Gaps = 8/753 (1%)

5 Query: 5 KKLKKIFLDYVIHIDRRSSPQKNRRVQNMILLTIFLFFIIFINFLVIGTDSKPGVNL 64
 KK +K LDFV+ RRR+P +NR RVGN+M+LTF+FFIFLINF+II+GTD KPGV+L
 Sbjct: 2 KKKQKIVLDYV--RRRGTVENKRVGVQNMILLTIFLFFIIFINFLVIGTDSKPGVNL 59

10 Query: 65 SKEAKVYQSSMTVQAKGTIYDRNGNPIARDATTYSLVAILISKNYTTATGQKLVQPSQ 124
 S+ AKRVYQ+++IT+QAKGTIYDRNG IA D+ITTS+YAI+ K++ +A+ +KLIVQPSQ
 Sbjct: 60 SGAQKRVQRTVTIQAQRGTIYDRNGTALAVDSTTYSYAILLOKSFVSASDEKLVQPSQ 119

15 Query: 125 YEKVASILENKLGMKNLVLQNLQKQLFQVSGSSGSLGYTTMADIKMTKESDIRGI 184
 YE VA IL+ LGMKK V+KGL +K LFCVSFG SSGS+SY+ M+ I+K ME + INGI
 Sbjct: 120 YETVADILKHLGMKKTIVIKQLKRGFLQVSGSGGSGISISTMTIQKAMEDAKINGI 179

20 Query: 185 GFSTSPRIYFNGIFASQFIQ-TLPQDDGDG-KLIVMTGLEAALNKVLSGTDGKVTYE 242
 F+TSPGR+YFNG FAS+FIG +L +D G K LVG TGLEA+ +K+LSS DG +TY+
 Sbjct: 180 APTTSPGRYFNGTYASEFGLASLTDDKKTGVKSLVGTGLEASFDKILSGQGVVITY 239

25 Query: 243 KDRSGWLLQTTATERRAVNGKDIYTTLSEPICTVLETQMDVFAEKTGKFPASATVNNAK 302
 KDR+G LLGT T +A++GRDIYTTLSEPICT LETQMDV F+ G+ ASATVNNAK
 Sbjct: 240 KDRNGTLLGTGKTVKKAIDGKDIYTTLSEPICTVLETQMDVFPQAKSNGQLASATVNNAK 299

30 Query: 303 TGEILATQRPPTYNPTSLKGVYDKNLOTNTYLLYDNFPEPSSMKVMTLASAIDSKHFN 362
 TGEILAT+ORPTYN TLG+ N Y+ L N FEPGSTMKVMTLA+AD K FN
 Sbjct: 300 TGEILATTQRPPTYNADTLKGLNTYKNVYSALHQN-FEPGSTMKVMTLAAIDKVFNP 358

35 Query: 363 TEVYNSAQ-YKIDALIIDMDVNEGLSSGSYMTFPQGFHSSNVGMVTLQKMDRDKWLN 421
 E +A+A IADA I+M +NEG+S+G YM + QGFA SNNVM LEQMG KW+N
 Sbjct: 359 NETFBNANGTLIADATIQWINSINEGISTQYMYAQGFAP+SNVGMVTLQKMDRDKWLN 418

40 Query: 422 YLSKFKGYPTFRPMHSSGLFPSDNEVTIAMSSFGQIGIVTQWMLRAFTSINDGMV 481
 YL+K+FG+PTFRG+ E G+PESDN VT AMS+PQGI VTQ+QMLRAFT+ISN+G M
 Sbjct: 419 YLTFRFGPYPTFRGLKDEAGI PPSDNVTQAMSAFQGISVITQIMLEAFTAISNNGEM 478

45 Query: 482 LQPFQISSIYDPMGTSTARKVVGKPVSKAASKTRDVMVTGTDPPYGTILYA-AGAF 540
 LQ+PQIS IYDPMT + RTA KE+VGKPVSK+AA+TR YM+ VGTD P +GLY+ P
 Sbjct: 479 LEPQFISQIYDPMTASFTANKKEIVGKPVSKAASBETQYMIWGTDPERGTLISKTGFP 538

50 Query: 541 VIOVGNSVAVKSGTAQIAQGGGGYLQ-GKNDTINSVVMVPSNEDFMVYTIQOPEK 599
 +I+VG+ VAVKSGTAQI E G Y G + +SVVMV++ PDF+MYVT+ +P+
 Sbjct: 539 IIKVGLFVAVKSGTAQIGSEDSGYQDGLTNYVYSVVMVFADEKDFLMVYTMKEPCH 598

55 Query: 600 FSITFWKDVVNPVLEQATMKEITLKPGLNDSHOCTYKLSKIVGNPHGVABELRNVL 659
 F FW+DVVNPVLEA M+T+ KP ++D+ QT YKL VG+NGP + ELERNVL
 Sbjct: 599 FGPLEWQDVVNPVLEAYLMQDTLTKPVVSDANRQTTYKLBNFVGNPGETSSELARNVL 658

60 Query: 660 QPIILNGSKSVKSKRPGANLAENEQLLVLTINKLETFPMYGSKANVQPAWNTGIVK 719
 QP++LG GSK+ KVS +PG L EN-Q+L+L++ E+PDYGM+K+NV+ FAKWTCI +
 Sbjct: 659 QPVVLGTSGSKIKKVSHPQGLTBTQGVLLSDDRFEVPEPMYGTSKNVTFKWTGIDI 718

70 Query: 720 TYXGTSGRVKRQSIDVWGSINKIKKIKITIGD 752
 ++KG+ SG+V EGS+DVGKS+ KIKK+ IT+GD
 Sbjct: 719 SFKGTDSGRVMKQSVDVGKSLKRIKKMTITLGD 751

A related GBS gene <SEQ ID 8691> and protein <SEQ ID 8692> were also identified. Analysis of this protein sequence reveals the following:

60 Lipop: Possible site: -1 Crand: 8
 McG: Discrim Score: -4.31
 GvH: Signal Score (-7.5): -7.07
 Possible site: 47
 >>> Seems to have no N-terminal signal sequence
 65 ALOM program count: 1 value: -13.90 threshold: 0.0

-1046-

INTEGRAL Likelihood = -13.90 Transmembrane 37 - 53 (30 - 60)
 PERIPHERAL Likelihood = 5.30 450
 modified ALOM score: 3.28

5 *** Reasoning Step: 3

----- Final Results -----

10 bacterial membrane --- Certainty=0.6562 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

15 ORF00411(301 - 2556 of 2856)
 Gp[677911|emb|CB870457.1][A94911(1 - 752 of 752) unnamed protein product {unidentified},
 homology to penicillin-binding protein 2x (S. pneumoniae)
 %Match = 77.4
 %Identity = 99.7 %Similarity = 99.9
 Matches = 750 Mismatches = 1 Conservative Sub.s = 1

20 66 96 126 156 186 216 246 276
 REKAFYGAIVITAIIMAVGIIYIQSNLSQVKEVNLKSKINKKQTEFDNAKQEVNELSNRDRITKIKADAGLTIQNDN

25 306 336 366 396 426 456 486 516
 IYRKVD*SVTFKKLKKIFLDYVVIHIDRRSPQKQKRVQNMILTIPLFFIFILNFIIVGTDSKFGVNLKKAQKVY
 VTFKKLKKIFLDYVVIHIDRRSPQKQKRVQNMILTIPLFFIFILNFIIVGTDSKFGVNLKKAQKVY
 10 20 30 40 50 60 70

30 546 576 606 636 666 696 726 756
 QQSMIVQAKSGIYDNRGNPIAEDATTSLYALISKNYITATGQKLYVQPSQYKVASILENKLGMKMLVLAQNLKGL
 QQSMIVQAKSGIYDNRGNPIAEDATTSLYALISKNYITATGQKLYVQPSQYKVASILENKLGMKMLVLAQNLKGL
 90 100 110 120 130 140 150

35 786 816 846 876 906 936 966 996
 FQVSPGSSGGIYDNRGNPIAEDATTSLYALISKNYITATGQKLYVQPSQYKVASILENKLGMKMLVLAQNLKGL
 FQVSPGSSGGIYDNRGNPIAEDATTSLYALISKNYITATGQKLYVQPSQYKVASILENKLGMKMLVLAQNLKGL
 170 180 190 200 210 220 230

40 1026 1056 1086 1116 1146 1176 1206 1236
 SGTDGKVTYKDRSGVNLGTATERRAVNGKDIYTLSEPICTVLETOMVFAKTKGKFASATVNVNAKTEILATSOR
 SGTDGKVTYKDRSGVNLGTATERRAVNGKDIYTLSEPICTVLETOMVFAKTKGKFASATVNVNAKTEILATSOR

45 250 260 270 280 290 300 310
 1266 1296 1326 1356 1386 1416 1446 1476
 PTYNPSTLKGVDKKNIGTYNTLLYDNFFPGSTWKMVLAISDKIIPNSTEVNSAQYKIADAVIRUDVNEGLSSGSY
 PTYNPSTLKGVDKKNIGTYNTLLYDNFFPGSTWKMVLAISDKIIPNSTEVNSAQYKIADAVIRUDVNEGLSSGSY

50 330 340 350 360 370 380 390
 1506 1536 1566 1596 1626 1656 1686 1716
 MTFPQGFHSSIVNGVILEQKMGKDKMLNLYSKFKPGYPTFRQMLHESGGLFPSDNEVTIAMSSPGQIGVTQVQMLRAF
 MTFPQGFHSSIVNGVILEQKMGKDKMLNLYSKFKPGYPTFRQMLHESGGLFPSDNEVTIAMSSPGQIGVTQVQMLRAF

55 410 420 430 440 450 460 470
 1746 1776 1806 1836 1866 1896 1926 1956
 TSISNDGVMVLQPPFISIIYDPNTGSTRARKEVVGKPKVSKAASKTRDYMVIVGTDPYVGTLYAAGAPVIVQVNDQSVYK
 TSISNDGVMVLQPPFISIIYDPNTGSTRARKEVVGKPKVSKAASKTRDYMVIVGTDPYVGTLYAAGAPVIVQVNDQSVYK

60 490 500 510 520 530 540 550
 1986 2016 2046 2076 2106 2136 2166 2196
 SGTQAQIAGGGGGLVQKNDTINSVAVPSENPOPIYVTVIQQEKFSITFWDDVNVNLEQATANNKTEILKPLNDESE
 SGTQAQIAGGGGGLVQKNDTINSVAVPSENPOPIYVTVIQQEKFSITFWDDVNVNLEQATANNKTEILKPLNDESE

-1047-

```

SGTACLAEQGGGGVLAQKQNTINSVVMVPSSENPDFIMYVITQPEKFSITFWKDVVNFLEQATAMKETTILKLPVINDSE
      570      580      590      600      610      620      630
5  2226      2256      2286      2316      2346      2376      2406      2436
   HQTGYKLGKIVGENPGHVAEELRRNLVQPIILGNGSKVSKVKRPGANLARNBQLLVLTINKLTLELPMYGWSKANVEQFA
   |||||
   HQTGYKLGKIVGENPGHVAEELRRNLVQPIILGNGSKVSKVKRPGANLARNBQLLVLTINKLTLELPMYGWSKANVEQFA
      650      660      670      680      690      700      710
10 2466      2496      2526      2556      2586      2616      2646      2676
   KWTGIKVTYGGSTSGKVRKQSIDVGKSIINKIKIKITIGD*HVFKYNGRCHSICPDSYCHSALH*VLPIEENWRATNA*R
   |||||
   KWTGIKVTYGGSTSGKVRKQSIDVGKSIINKIKIKITIGD
      730      740      750

```

15 SEQ ID 8692 (GBS352d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 15 & 16; MW 105.5kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 17 & 18; MW 80.5kDa), in Figure 182 (lane 3; MW 80kDa) and in Figure 185 (lane 4; MW 105kDa). Purified GBS352d-GST is shown in lane 5 of Figure 236.

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 954

A DNA sequence (GBSx1012) was identified in *S.agalactiae* <SEQ ID 2907> which encodes the amino acid sequence <SEQ ID 2908>. Analysis of this protein sequence reveals the following:

```

25  Possible site: 36
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
30  bacterial cytoplasm --- Certainty=0.1950 (Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 955

A DNA sequence (GBSx1013) was identified in *S.agalactiae* <SEQ ID 2909> which encodes the amino acid sequence <SEQ ID 2910>. This protein is predicted to be unnamed protein product (mraY). Analysis of this protein sequence reveals the following:

```

40  Possible site: 18
   >>> Seems to have a cleavable N-term signal seq.

   INTEGRAL  Likelihood =-15.12  Transmembrane  56 - 72 ( 47 - 76)
   INTEGRAL  Likelihood =-14.70  Transmembrane  203 - 219 ( 198 - 223)
45  INTEGRAL  Likelihood =-6.69   Transmembrane  318 - 334 ( 315 - 335)
   INTEGRAL  Likelihood =-6.64   Transmembrane  83 - 99 ( 79 - 103)
   INTEGRAL  Likelihood =-5.52   Transmembrane  179 - 195 ( 175 - 197)
   INTEGRAL  Likelihood =-5.31   Transmembrane  232 - 248 ( 230 - 249)
   INTEGRAL  Likelihood =-3.08   Transmembrane  119 - 135 ( 119 - 137)
50  INTEGRAL  Likelihood =-2.87   Transmembrane  151 - 167 ( 147 - 167)
   INTEGRAL  Likelihood =-2.34   Transmembrane  254 - 270 ( 254 - 270)

```

-1048-

----- Final Results -----

bacterial membrane --- Certainty=0.7050(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2911> which encodes the amino acid sequence <SEQ ID 2912>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.55 Transmembrane 52 - 68 (48 - 75)
 INTEGRAL Likelihood = -9.39 Transmembrane 175 - 191 (171 - 194)
 INTEGRAL Likelihood = -8.12 Transmembrane 30 - 46 (23 - 48)
 INTEGRAL Likelihood = -6.37 Transmembrane 121 - 137 (119 - 145)
 INTEGRAL Likelihood = -6.32 Transmembrane 293 - 309 (287 - 309)
 INTEGRAL Likelihood = -5.31 Transmembrane 204 - 220 (202 - 221)
 INTEGRAL Likelihood = -5.20 Transmembrane 151 - 167 (150 - 170)
 INTEGRAL Likelihood = -4.67 Transmembrane 226 - 242 (224 - 244)
 INTEGRAL Likelihood = -0.11 Transmembrane 91 - 107 (91 - 107)

----- Final Results -----
 bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP: CAB70458 GS: A94911 unnamed protein product [unidentified]
 Identities = 244/309 (78%), Positives = 273/309 (87%), Gaps = 1/309 (0%)

Query: 1 LKKIGGQCMHEDVKQHLAKAGTPTMGTVFLVATVATVSLVSLF-SIKNTQSLALISGIL 59
 LKKIGGQCMHEDVKQHLAKAGTPTMGTVFL+VA VSL+ S+ S+N+L+L GIL
 Sbjct: 26 LKKIGGQCMHEDVKQHLAKAGTPTMGTVFLVALLVSLIPSIILSKNSGILGATPGIL 87

Query: 60 SIVVYIGIIFGLDDFLKIKFQINBGLTAKQKIALQLVGGLMFFYFLHVSFSGISSINVFY 119
 S+V+TYGIIGFLDDFLKIKFQINBGLT KQK++LQL+ GL+FYF+HV PSF S+IN+FG+
 Sbjct: 86 SVVLYIGIIFGLDDFLKIKFQINBGLTPKQKMSLQIAGLIFFYPVHVLPSGTSAINIFGF 147

Query: 120 QLPGLIFLYFFVLFWVGFPSNAVNLTDGIDGLASISVVISLITYGIAYVQSQDFVLLI 179
 L +G Y FFLVFWVGFPSNAVNLTDGIDGLASISVVISL+TYG+IAY Q+QD+LH+I
 Sbjct: 148 NLEVGVLVYAFFVLFWVGFPSNAVNLTDGIDGLASISVVISLITYGIAYNQDPTDILLI 207

Query: 180 GAMIGALLGFFFCFPHKPAKVPFMDVGSALGAMLAISIALRQSWTLLIIGIVVLTSS 239
 MIGALLGFF FPHKPAKVPFMDVGSALGAMLAISIALRQSWTLL IG VYV ETSS
 Sbjct: 208 VIMIGALLGFFVPHKPAKVPFMDVGSALGAMLAISIALRQSWTLLIFGIVVFTSS 267

Query: 240 VMLQVSYPKYTKKKYGGERRIFRMTPPHHHLELGGSGKGNKSEWQVDAFLMGVSLAS 299
 VMLQV+YFKYTKK G G+RIFRMTPPHHHLELGG+SGK GNSEW+VDAFLM +G S
 Sbjct: 268 VMLQVAYPKYTKKKYGGERRIFRMTPPHHHLELGGSGKGNKSEWQVDAFLMIGIFMS 327

Query: 300 LLLVLAHLV 308
 + LAILY+
 Sbjct: 328 AITLAILYL 336

An alignment of the GAS and GBS proteins is shown below.

Identities = 244/309 (78%), Positives = 273/309 (87%), Gaps = 1/309 (0%)

Query: 28 LKKIGGQCMHEDVKQHLAKAGTPTMGTVFLVALLVSLIPSIIILSKNSGILGATPGIL 87
 LKKIGGQCMHEDVKQHLAKAGTPTMGTVFL+VA VSL+ S+ S+N+L+L GIL
 Sbjct: 1 LKKIGGQCMHEDVKQHLAKAGTPTMGTVFLVATVATVSLVSLF-SIKNTQSLALISGIL 59

Query: 88 SVVLYIGIIFGLDDFLKIKFQINBGLTPKQKMSLQIAGLIFFYPVHVLPSGTSAINIFGF 147
 S+V+TYGIIGFLDDFLKIKFQINBGLT KQK++LQL+ GL+FYF+HV PSF S+IN+FG+
 Sbjct: 60 SIVVYIGIIFGLDDFLKIKFQINBGLTAKQKIALQLVGGLMFFYFLHVSFSGISSINVFY 119

-1049-

Query: 148 YLEVGVLVAFVFLFWVGFSSNAVNLTDGIDGLASISVVISLITYGILAYNQTFDILLII 207
 L +G Y PFLVFWVGFSSNAVNLTDGIDGLASISVVISL+TYG+LAY Q-QPD+LI+I
 Sbjct: 120 QLPLGIFLFFFLFWVGFSSNAVNLTDGIDGLASISVVISLVTYGIVAYVQSQPDVLLLI 179

5 Query: 208 VIMIGALLGFYVFNHHPAKVFMKDVGSALGAMLAASIALRQEWTLIFGFPVYVFETSS 267
 MIGALLGFF FNHHPAKVFMKDVGSALGAMLAASIALRQEWTL IG VIV ETSS
 Sbjct: 180 GAMIGALLGFPFNHHPAKVFMKDVGSALGAMLAASIALRQEWTLIGIVIVVFETSS 239

10 Query: 268 VMLQVAYPKYTKKKTGVGKRIFRMTFPHHILELGGVSGKGNWSEMKVDAPLWAGIETSS 327
 VMLQV+YKPYTKKK G G+RIFMTFPHHILELGG+SGKG KWSN+VDAFLW +G S
 Sbjct: 240 VMLQVSYPKYTKKYGGRIRFMTFPHHILELGGSGKGKWSNQVDAFLWGVSLAS 299

Query: 328 AITLAILYL 336
 + LAILY+
 15 Sbjct: 300 LLVLAILELV 308

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 956

20 A DNA sequence (GBSx1014) was identified in *S. agalactiae* <SEQ ID 2913> which encodes the amino acid sequence <SEQ ID 2914>. This protein is predicted to be autoaggregation-mediating protein (dead). Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3018 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14444 GB:Z99116 similar to ATP-dependent RNA helicase
 [Bacillus subtilis]
 Identities = 215/436 (49%), Positives = 310/436 (70%), Gaps = 5/436 (1%)

35 Query: 3 FKFNFKPKYIQRALDELKFDPTDVQAKLIPVVRSGRDLVGSKTGSKGHTHFLPIFEK 62
 F+ + KP+I A+ L F +PTD+Q +LIP V ++G+S+TG+GRTH +LLP+ K
 Sbjct: 6 FELVELKPTFLDAVHRLGFYEPTDIQGLIPAVLAKESVIGQSQTGTGKTHAYLLPLINK 65

40 Query: 63 LDSESDDVQVVITAPSGRELGTQIYQATKQIAHSE-QEIRVNVYVGTGDKLRQIEKLKVS 121
 +D + D VQVVITAP+REL QIYQ +I + E +IR ++GGTDK +Y+KLK+
 Sbjct: 66 IDPAKDVQVVITAPTRERLANQIYQKALKTQCEGSGQIRSKCFPGTGDQKSIDKLKI- 124

45 Query: 122 QPHVIGTIFGRYIDLKSGDLAIHKAHTFVVDADMTLDMGFLITVDVKIAGSLPKDVQIL 181
 QPH+V+GTGRI DL+K L+HKA +V+DEAD+ LDMGFL VD I +P+D+Q+L
 Sbjct: 125 QPHLVGTIFGRYIDLKSGDLAIHKAHTFVVDADMTLDMGFLITVDVKIAGSLPKDVQIL 184

Query: 182 VFSATIPKQLPFLKYLTFNFMKIKITATVIADTFIDNWLSTKGRDNQAQILELSKIMQ 241
 VFSATIP+KL+PFLKYL+ NP ++ V A I++ L+ +K RDK+ + ++ +
 50 Sbjct: 185 VFSATIPKELKPLKLYMENPKYAHVSPQVTAARIKHILIPSKHRDKKLLFOIMSHIN 244

Query: 242 PYLAMIVPNTKERADIELSHYSSNGLVAKIHOGIAPRERKRNQVKNLSPEYIVATDL 301
 PYL ++F NTK AD + YL+ G+K+ +HGG+ PRERK+M Q+ +LEF YI+ATDL
 55 Sbjct: 245 PYLGIVFANTKNADHIAQYLTGKMKIGLHAGGLTPRERKKNWQKQINDLSFTYIATDL 304

Query: 302 AARGIDIEGVSHVINDAIPDLSFFVIRVGTGRNGLSGTATILYQPSDDSDIRELEKLG 361
 AARGIDI+GVSHVIN +P DL P+VIRVGTGRNGLSGTATILYQPSDDSDIRELEKLG
 Sbjct: 305 AARGIDIEGVSHVINELPDLDYFVIRVGTGRNGLSGTATILYQPSDDSDIRELEKLG 364

60 Query: 362 INFIPKVIKNGEFOQTYDRDRNRNRSYQKLDTEMIGLVKKKKKKIKPGYKKIQMKVD 421
 I F ++ GSt+ DR RR R+K+ +D E+ + KK KK+KPGYKKK+ ++++

-1050-

Sbjct: 365 IEFYLELEKGEWKKGGDDORRRKKKKTPNEAD-ETAEHLVKKPKVKPGYKKKMSYEME 423

Query: 422 EKRRKERRASNRANKR 437

5 Sbjct: 424 KIKKKQRR--NQSKKR 437

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2915> which encodes the amino acid sequence <SEQ ID 2916>. Analysis of this protein sequence reveals the following:

10 Possible site: 54
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2315(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 382/447 (85%), Positives = 420/447 (93%)

20 Query: 1 MSFIDNFNKFYIQRAIDELKLFVDPDVGAKLIPVVRSGRDLVGSBKTGSGKTHIFLLPIF 60
 MSFKD++FK Y+Q+AL+E+ FV+PT+VQ +LIP+V SGRDLVGSBKTGSGKTHIFLLPIF
 Sbjct: 1 MSFQDYHFQYVQALBEEIGFVNPTFVQKRLIFIVNSGRDLVGSBKTGSGKTHIFLLPIF 60

25 Query: 61 EKLDESSDDVQVITAPSRELGTQIYQATQIAEHSQEI RVNVTGQTDKLRQIEKLKV 120
 EKLDE+ +VQVITAPSREL GTQI A IQIA+H ++EIR+ RTVGQTDKLRQIEKLK
 Sbjct: 61 EKLDEAKAEVQVITAPSRELATQIFDACKQIAGHFQEIRLAHTVGGTDKLRQIEKLKD 120

30 Query: 121 SQPHIVIGTSGRIYDLVKSGLAIHKARFVVDDEADMTLWGFLDVTOKIAGSLPKVDQI 180
 SQPHIVIGTSGRIYDLVKSGLAIHKA TFVVDDEADMT+DMSFLDVTOKIA SLPK VOI
 Sbjct: 121 SQPHIVIGTSGRIYDLVKSGLAIHKATTFVVDDEADMTMW3FLDVTOKIAPSLPKSVQI 180

35 Query: 181 LVFSATIPQKQLFLEKKYLINPVEKIKATFVIADTTIDNMLSTKGRDKNAQIIELESLM 240
 LVFSATIPQKQLFLEKKYLINFPV+E+IKT TVIADTTIDNML+STKGRDKN Q+LE+ K M
 Sbjct: 181 LVFSATIPQKQLFLEKKYLINPVEIQIKTKTVIADTTIDNMLVSTKGRDKNQLLEILKTM 240

40 Query: 241 QPYLAMIFVNTKERADDELH+SYLSSNGLKVAKIHGGIAPRKRKIMNQVNHLEFEYIVATD 300
 QPY+AM+FVNTKERAD+LH++L++NGLKVAKIHGGI PRKRKIMNQV L+FEYIVATD
 Sbjct: 241 QPYMAMLFVNTKERADDLHAFLTANGKVAKIHGGIPRKRKIMNQVKKLDPEYIVATD 300

45 Query: 301 LAARGIDIBGVSHVINDAIPQDLSPFVHRVGRTRNGLSGTAITLYOPSDSDSIRELEKLI 360
 LAARGIDIBGVSHVINDAIPQDLSPFVHRVGRTRNG++GTAITLYOPSDSDS+ILEK+
 Sbjct: 301 LAARGIDIBGVSHVINDAIPQDLSPFVHRVGRTRNGMAGTAITLYOPSDSDS+KLEKLM 360

45 Query: 361 GINFIKPVIKNGEFQDTYDRDRNRREKSYQKLDTEMIGLVKGGKKKKIPGYKKKIQKV 420
 GI F PKV+KNGEFQDTYDRDR NRK+YQKLDTEMIGLVKGGKKKK+KPGYKKKIQN V
 Sbjct: 361 GIATFPKVLKNGEFQDTYDRDRNRREKAYQKLDTEMIGLVKGGKKKKVKGPGYKKKIQNAV 420

50 Query: 421 DEKRRKERRASNRANKRABERKAKKQSF 447
 DEKRRKERRA NRANKRABERKAKKQ F
 Sbjct: 421 DEKRRKERRASNRANKRABERKAKKQHP 447

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 957

55 A DNA sequence (GBSx1015) was identified in *S.agalactiae* <SEQ ID 2917> which encodes the amino acid sequence <SEQ ID 2918>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have a cleavable N-term signal seq.

-1052-

230 240 250 260

SEQ ID 8694 (GBS8) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 2 (lane 5; MW 31kDa), Figure 63 (lane 2; MW 31.3kDa), Figure 66 (lane 2 & 3; MW 31kDa), in Figure 178 (lane 2; MW 31kDa), in Figure 179 (lane 3 & 4; MW 31kDa) and in Figure 180 (lane 3; MW 31kDa). It was also expressed in *E. coli* as a GST-fusion product, with SDS-PAGE shown in Figure 66 (lanes 4 & 5; MW 56kDa) and in Figure 180 (lanes 4 & 5; MW 55kDa).

GBS8-His was purified as shown in Figures 189 (lane 7), 211 (lane 3), 228 (lanes 4-5) and 230 (lanes 3-6). Purified GBS8-GST is shown in Figure 209, lane 6.

The GBS8-His fusion product was purified (Figure 90A) and used to immunise mice (lane 2 product; 12.9µg/mouse). The resulting antiserum was used for Western blot (Figure 90B), FACS (Figure 90C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 958

A DNA sequence (GBSx1016) was identified in *S. agalactiae* <SEQ ID 2921> which encodes the amino acid sequence <SEQ ID 2922>. Analysis of this protein sequence reveals the following:

```

Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3991(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 959

A DNA sequence (GBSx1017) was identified in *S. agalactiae* <SEQ ID 2923> which encodes the amino acid sequence <SEQ ID 2924>. This protein is predicted to be probable amino-acid abc transporter permease protein in *idh-deor* inter. Analysis of this protein sequence reveals the following:

```

Possible site: 56
>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood =-11.62 Transmembrane 50 - 66 ( 41 - 74)
INTEGRAL Likelihood = -0.90 Transmembrane 226 - 242 ( 226 - 242)
INTEGRAL Likelihood = -0.53 Transmembrane 80 - 96 ( 80 - 96)

----- Final Results -----
bacterial membrane --- Certainty=0.5649(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

-1053-

>GP:CAB15985 GB:X99124 similar to amino acid ABC transporter
(permease) [Bacillus subtilis]
Identities = 90/224 (40%), Positives = 137/224 (60%), Gaps = 10/224 (4%)

5 Query: 28 NKAVLDRAIPISILERLPITLLTVAGALFGLIALIFAVVKINRVKILYPIQALFVSFLRG 87
W+ + A P++++ LPITL + +A +F +I LI A++ N++ +L+ + L++SF RG
Sbjct: 6 NEFMISAPFTLLIQALPITLFMAIAAMFIAIIGGLILALITKNKIPVLHQLSKLIVISFPRG 65

10 Query: 88 TPILVQMLSYGIGLPLFLKFLNQKYGPDNNAINALPASVFAITAFAPNEAAYTSETIRAAI 147
P LQVL L YYG+P + + + +A AI + +A +E RAA+
Sbjct: 66 VPTLVQFLYLYGFLQPLFPMESK-----MTALTAIIQLSLNNAAYLARIFFRAAL 115

Query: 148 LSVQGEIEAARS LGMTSAQVYRRVILIPNAVVATPTLINTLIGLTKGTSIAFNAGIV 207
SVD G++EA S+GMT Q YER+I+P A A P NT IGL K TSLAF G++EM
15 Sbjct: 116 NSVDDQGLEACLSVGMTKFQAYRRIILPQAIRNAIPATGNTPIGLTKETSIAFTLVGMEN 175

Query: 208 FQAQIMGSDYRYFERYISVALVYVAWSFLIEQLGNALSKMA 251
FAQ ++ + +YFE Y++VA+VYM ++ + L + ER M+
Sbjct: 176 FQAQIMGSDYRYFERYISVALVYVAWSFLIEQLGFLERAMS 219

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2925> which encodes the amino acid sequence <SEQ ID 2926>. Analysis of this protein sequence reveals the following:

Possible site: 43
>>> Seems to have no N-terminal signal sequence

25 INTEGRAL Likelihood = -7.27 Transmembrane 80 - 96 (74 - 104)
INTEGRAL Likelihood = -1.06 Transmembrane 207 - 223 (207 - 223)
INTEGRAL Likelihood = -0.90 Transmembrane 110 - 126 (110 - 126)

----- Final Results -----
bacterial membrane --- Certainty=0.3909 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9167> which encodes the amino acid sequence <SEQ ID 9168>. Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

40 INTEGRAL Likelihood = -7.27 Transmembrane 50 - 66 (44 - 74)
INTEGRAL Likelihood = -1.06 Transmembrane 177 - 193 (177 - 193)
INTEGRAL Likelihood = -0.90 Transmembrane 80 - 96 (80 - 96)

----- Final Results -----
bacterial membrane --- Certainty=0.391 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
45 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 212/267 (79%), Positives = 238/267 (88%)

50 Query: 1 MNQFII TGGNSWYNLVSQVFAKGLPSWKAVLDRAIPISILERLPITLLTVAGALFGLILA 60
M LT GW++Y+ L+S +P GKLPW AV DAIP+I++RLPITL LT++GA FGL+IA
Sbjct: 31 MTSVLTSGWAFYDYLTISP IPHKGLPSWHAIFDAIPNI IQRLPITLGLTSGCATFGLVLA 90

55 Query: 61 LIFAVVKINRVKILYPIQALFVSFLRGTPILVQMLSYGIGLPLFLKFLNQKYGPDNNINA 120
LIFA+VKIN+VK+LYPIQA+ FVSFLRGTPILVQML+YYGIGLPLFLKFLNQKYGPDNN+NA
Sbjct: 91 LIFALVKINRVKILYPIQALFVSFLRGTPILVQMLITTYGIGLPLFLKFLNQKYGPDNNVNA 150

Query: 121 IPASVFAITAFAPNEAAYTSETIRAAISVDQGEIEAARS LGMTSAQVYRRVILIPNAVV 180
IPAS+FAITAFAPNEAAY TSETIRAAISVD GEIEAA+SLGMS QVYRRVILIPNA VV
60 Sbjct: 151 IPASFAITAFAPNEAAYTSETIRAAISVDITGEIEAARS LGMTSAQVYRRVILIPNAVTV 210

Query: 181 ATPILINTLIGLTKGTSIAFNAGIVEMFAQAQIMGSDYRYFERYISVALVYVAWSFLIE 240
A PTILN LIGLTKGTSIAFNAGIVEMFAQAQI+GGSDYRYFERYISVALVYVAWS+L+E

-1054-

Sbjct: 211 AIPTILINGLIGLTKGTSLAFNAGIVMFAQAQILGSDYRYFERYISVALVYWSISLIME 270

Query: 241 QLGNAIERKQAIKAPRHLDTEIPQGV 267

Q+G IE KNAIKAP +E G +R

Sbjct: 271 QVGRILNIEKQAIKAPKQARNEKLGELR 297

There is also homology to SEQ ID 4794.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 960

A DNA sequence (GBSx1018) was identified in *S. agalactiae* <SEQ ID 2927> which encodes the amino acid sequence <SEQ ID 2928>. This protein is predicted to be amino acid ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3205 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA000329 GB:AF008220 putative amino acid transporter [Bacillus subtilis]
Identities = 121/247 (49%), Positives = 176/247 (70%)

Query: 1 MIKLRQLTKSPGQKVLKLDLIEKGQVVALVGASGAGKSTFLRSMNYLESFDYGTIEI 60

MI+++ + K F VL ++L + KG+VV ++G SG+GK+TFLR +N LE PD G I I

Sbjct: 1 MIEILNIHQPGIHNVLKGINLTVRKGEVVTIIGPSGSGKTTFLRCLNLERPDEGIISI 60

Query: 61 DDFKVPFSGISKDILLRKKLAMVFPQNFERRRTALDNVKEGLKIVKMSDQEAETRIA 120

D ++ + SK ++ LR++ AMVFQQ++LF +T ++NV EGL I +KM Q+A +A

Sbjct: 61 HDKVINCRFPSSKEVHWLRKQZAMVFQCYHLEAKHTVLENVMEGLTIARKMRKQDAYAVA 120

Query: 121 RDELAKVGLADREKYYPHLSGGQKQVALARALAMKPDVLLDDEPTSLDPELVGGEVK 180

+EL KVGL D+ YP LSGGQKQRV +ARALA+ PDVLL DEPT+ALDPELVGGEV +

Sbjct: 121 ENELRKYVGLQDKLNAYFQQLSGGQKQVGIARALAIHPDVLLDDEPTSLDPELVGGEV 180

Query: 181 SIADAAKQQQTMVLVSHDMNFPYQVADKVLFLKGRILLESCTPEQLFNHPLEERTKEFFA 240

+ + K G TM++V+H+M F +V+D+V+V++G I+E GTPR++F H ++RT++F

Sbjct: 181 VMLEIVTKGATMIVVTHEMEFARRSDQVFMDEGVIVQGTPEFVFRHTTKDKRTQPLR 240

Query: 241 SYNKSYL 247

+ YL

Sbjct: 241 RVSPFYL 247

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2929> which encodes the amino acid sequence <SEQ ID 2930>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1840 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 199/247 (80%), Positives = 229/247 (92%)

-1055-

Query: 1 MIKLPQLTKSPGQKVLKDLIDIEKGQVVALVGASGAGKSTPLRSNYLSEPDYGTIRI 60
MI +R L+K+PBGQKVL D L LDIKEGQV+ALVGASGAGKSTPLRS+NYL+PD G+I I
5 Sbjct: 2 MITIRNLSTKSPGQKVLSDIALDIEKGQVIALVGASGAGKSTPLRSNYLSEPDYGTIRI 61

Query: 61 DDFKVPKISISKDILTLRRKLAMVFPQNLFRRTALDNVKEGLK+VKK+SDQKATRIA 120
DF VDF++I+ + +L LRRKLAMVFPQNLFRRTAL+NVKGLK+VKK+SDQKAT+AA
Sbjct: 62 GDFTVDFETITTEQVILIRKLAMVFPQNLFRRTALDNVKEGLK+VKK+SDQKATKLA 121

10 Query: 121 RDLKLVGLADREKYPVPHLSGGQKQVVALARALAMKPDVILLDEPTSLDPELVGEVEK 180
+ ELAKVGLADR+ +YRHLSGGQKQVVALARALAMKPDVILLDEPTSLDPELVGEVEK
Sbjct: 122 QAEKLVGLADRHGHYPRHLSGGQKQVVALARALAMKPDVILLDEPTSLDPELVGEVEK 181

15 Query: 181 SIADRAKSGQTMVLVSHDMNPFYQVADKVLFLKGRILESGTPEQLFNHPLERTKEPFA 240
SI DAAK GQTMVLVSHDMNPFYQVAD+VLF++G+ILE GTPE++F HP +ERTKEPFA
Sbjct: 182 SITDAKSGQTMVLVSHDMNPFYQVADKVLFLDQKILRQSTPEVFRHPQKERTKEPFA 241

Query: 241 SYNKSYL 247
SY+K+Y+
20 Sbjct: 242 SYSKTYI 248

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 961

25 A DNA sequence (GBSx1019) was identified in *S. agalactiae* <SEQ ID 2931> which encodes the amino acid sequence <SEQ ID 2932>. Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0831 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

35 The protein has homology with the following sequences in the GENPEPT database.

>GP:BA807290 GB:AP001519 thioredoxin reductase (NADPH) [Bacillus halodurans]
Identities = 173/302 (57%), Positives = 234/302 (77%)

40 Query: 1 MYDTLIIGSGPGMTAALYAARSNLKVLIEQAGPGQMNITARIENYPGYDHISGPGLS 60
+YD +I G+GP GMTAA+Y +R+NL ++E+G PGGQM NT ++ENYPG+DHI GPGLS
Sbjct: 7 VYDVVIAGAGPAGMTAAVYTSRANLSTVMVERGVPGQMNITEDVENYPGDFHILGPGLS 66

Query: 61 MKMYEPLKPFZVEHLYGIVQRVENDGQKRVITRDSYEAKTVILATGAKNSLLGVPGEE 120
KQ+E +KF E+ YG ++ + + CD+K V ++ Y+A+ VI+ATCA+ LGVPGE+
45 Sbjct: 67 TKVPEHAKKPGAEYAYGDIKETIDQDLKLVKAGNKRYKARAVIVATGAEYKLLGVPGEK 126

Query: 121 EYTSRGVSYCAVCDGAPFRDQDLLVVGCGDSAVEAVFLTPQAKSVTI IHRDQLRAQKV 180
E + RGVSYCAVCDGAPF+ ++L+VVGCGDSAVEAV+LT+FA VTI IHRDQLRAQK+
50 Sbjct: 127 ELGSRGVSYCAVCDGAPFPGKELVVGCGDSAVEAVYLTRFASKVTI IHRDQLRAQKI 186

Query: 181 LQDRAFANKEKIFVMDSVVKEIKNSIKVSGVTVENLKTCGISEMTPGGVETIYVGLKPHS 240
LQ RAF M+KI+F+MD VVK+I G + KVS VT+E+ KIGE + GVTY+G+ P +
Sbjct: 187 LQDRAFNDKIEPIWDHVVQINZTDGVSSVTI SHAKTGRQDPKTDGVETIYVGLPLN 246

55 Query: 241 SMVSELGITDTEKWLVDLTNMTKTSIPGLVAGDVRQKDLRQIATVAGSALAGQVYNYI 300
V L I ++ G++T+ M+TS+PG+K GDVR+K LRQI TA G+G+AA Q V +YI
Sbjct: 247 EAVKNIILNIDBSYIVTNEEMTSVPG+PAAGDVRKSLRQIVTATGDSGLAANQVQHYI 306

Query: 301 TE 302
E
60 Sbjct: 307 EE 308

-1056-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2933> which encodes the amino acid sequence <SEQ ID 2934>. Analysis of this protein sequence reveals the following:

```

Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0386 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 236/300 (78%), Positives = 273/300 (90%)

Query: 1  MYDTLIIGSGPGGNTAALYAARENLKVLIECGAPGGQNNNTAEIENYPGYDHISGPGLS 60
      MYDTLIIGSGP GNTAALYAARENL V +IECGAPGGQNNNT +IENYPGYDHISGPGL+
Sbjct: 1  MYDTLIIGSGPAGNTAALYAARENLVALLIECGAPGGQNNNTFDIENYPGYDHISGPGLA 60

Query: 61  MNMYEPLKFEVHEIYGVIVRVENDGDVKKVITEDSEYEAQTVILATGAKNELLGVPGEE 120
      MNMYEPLKFP VE+YGVIVQ++EN GD K V+TED SYEAQTVI+ATGAK +LGVPGEE
Sbjct: 61  MNMYEPLKFNVENIYGVIVQKIENTFGDYKCVLTEDASEYEAQTVIATGAKYRVLGVPGEE 120

Query: 121  EYTRSGVSYCAVCDGAFPRDQDLVAVGGGDSAVEBAVFLTQPAKSVTIIHRRDQLRACKV 180
      YTRSGVSYCAVCDGAFPRDQDLVAVGGGDSAVEBA++LTQPAK VT++HRRDQLRACK+
Sbjct: 121  YYTRSGVSYCAVCDGAFPRDQDLVAVGGGDSAVEBAIYLTQPAKVVTVHRRDQLRACKI 180

Query: 181  LQDRAPANEKIKFVWDSVVKELKGNKIKVSGVTVENLKTGEISEMTFGGVFIYVGLKPHS 240
      LQDRAPAN+K+ P+WDSVVKEL+GN+IKVS V +EN+KTG+++ PGGVFIYVG+ P +
Sbjct: 181  LQDRAPANDKVDVWDSVVKELQGNIDIKVSNVLIENVKTQVTDHAFGGVFIYVGNFV 240

Query: 241  SMVSELSGITDETGWVLTDTNMKTSIPGLYAGDVRQKDLRQITAVGGAIAAGQGVNYI 300
      MY +L ITD GW++TD ++M+TSIPG++AIGDVRQKDLRQI TAVG+GAIAAGQGVY+Y+
Sbjct: 241  GNVKDLKITDSRGWLTIDHMKTSIPGIPALGIDVRQKDLRQITAVGGAIAAGQGVHYL 300

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 962

A DNA sequence (GBSx1020) was identified in *S.agalactiae* <SEQ ID 2935> which encodes the amino acid sequence <SEQ ID 2936>. Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3626 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB15163 GB:299120 similar to nicotinate
      phosphoribosyl transferase [Bacillus subtilis]
Identities = 309/476 (64%), Positives = 384/476 (79%), Gaps = 2/476 (0%)

Query: 2  YKDSLSLTHTDLYQINM+QVYFNKGIHNKRAVPRAYFRKVPBNKYAVFAGIERIVRYLE 61
      +KDSLS+LHTDLYQINM + Y+ GIH K+A+FR +FR++PFENGVAVFAGLE+ + YLE
Sbjct: 6  FKDSLSLSLTHTDLYQINMAETVYKRGDIHKKKAIPLFRRLPFRNGYAVFAGLEKALEYLE 65

Query: 62  NLSFSDSDLSYLE-ELGYPEEFLDYLNKLMELTVKSABQDLVFAFNEFLVQIBGFLAQC 120
      N F+DSDLSYL+ ELGY E+P++YL+ L ++S KEG+LVF NKP++++E FL +
Sbjct: 66  NFKFTSDSDLSYLQDELGYHEDFTIEYLRGLSPFGSYLMKQEGELVFNNEPTMRVZAPLVEA 125

```

-1057-

Query: 121 QLVEITALNIINQTLVATKAARISVIEDEPLLEPGTTRAAQEMDAIIGWTRAAIIGGN 180
 QL-ETA-LNI-N-QTL-ATKAARIS- VI DE LEPTGTRA EMDAA-WG RAA-IG3 +
 Sbjct: 126 QLIEITALNIINQTLVATKAARIGVIGDEVALEPGTTRAAHEMDAAWGAARALIGGFS 185

5 Query: 181 ATSNVRAGKIFNIPVSGTHAHLVQTYGDDYQAFKAYAEHTKDCVFLNDYDTLRGVGVN 240
 ATSNVRAGK FNIIPVSGTHAHLVQ Y D-Y AFK YAEHTKDCVFLNDYDTLR G+VN
 Sbjct: 186 ATSNVRAGKRFNIPVSGTHAHLVQAYRDEYTAFFKAYAEHTKDCVFLNDYDTLRSG+VN 245

10 Query: 241 AIRVAKEMGKINFLGVRLSDGLAYLSKKVRQQLDDAGFPNAKIYASNDLENTIIMLK 300
 AIRVAKR G+IINP+G+RLDSGLAYLSKK R+ LD+AGF +AK+ AS+DLDE-TI+MLK
 Sbjct: 246 AIRVAKEPCDRINFIQRLDSGLAYLSKKARMLDEAGFTDAKVIASSDLENTIIMLK 305

Query: 301 MQAKIDVWGVTGLITAYDQFALGAVYKIVSIETDAGSMRDTIKLSNNAEKVSTPGKQ 360
 Q A-IDVWGVTGLITAYDQFALGAVYK+IE D G M DTK+S+N EKV+TPG+K+
 Sbjct: 306 AQGARIDVWGVTGLITAYDQFALGAVYKLVAIED-QGVDTTIKISSNPEKVTTPGRKK 364

15 Query: 361 VWRITSRANGKSGDYITFAIDVQLDIEIMFHTPTTYTIINKIVRDFDAVPLLDVDFPKG 420
 V+RI +++ SGDYI D V + MFP +T+I+K V +P A L IFL+KG
 Sbjct: 365 VYRIINQSNHSGDYIALYDEQVNDQKRLMFPHTVTFISKPVTFYAKDLHELIFPKG 424

20 Query: 421 KLWYQLFSLQEIQYGRKEPDGLWDEYKRVNFCQYVVDLARDWQNMOLIDRIR 476
 L YQ P + +IQ+Y + LM+EYKR+ P+YFVDL+ D W NM I ++
 Sbjct: 425 ILCYQNFESIDIQYVQDNLSLWEEYKRISKPEEYVVDLSDCSNMQRIHEVK 480

- 25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2937> which encodes the amino acid sequence <SEQ ID 2938>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3192 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 35 An alignment of the GAS and GBS proteins is shown below.

Identities = 409/464 (88%), Positives = 446/464 (96%)

Query: 1 MYKDSGLTHTDLVQINMQVFNKGIHNRKAVPEYFRKVPFNGYAVFAGLRIVRYL 60
 MYKDSGLTHTDLVQINMQVFN +GIHN+ AVFE YFRK PF NGYAVFAGLR+AV YL
 40 Sbjct: 1 MYKDSGLTHTDLVQINMQVFNKGIHNRKAVPEYFRKVPFNGYAVFAGLRIVRYL 60

Query: 61 ENLSFSDSLVLEELGYPPEFLDYLNKLMELTVKSACEGDLVFANSPVLQVIGPLAQ 120
 E PS+DL+VLEELGYPR FL YLK L+ELT++SAKRGDLVFNRP+VQ+BGPL QC
 45 Sbjct: 61 BQFQFSETDLVLEELGYPENFLTYLKLRLLELTIRSAKEDLVFNRPVQVIGPLAQ 120

Query: 121 QLVEITALNIINQTLVATKAARISVIEDEPLLEPGTTRAAQEMDAIIGWTRAAIIGGN 180
 QLVEITA-LNI-N-QTL-ATKAARISVIEDEPLLEPGTTRAAQ-DAIIGWTRAA-IGGA-
 Sbjct: 121 QLVEITALNIINQTLVATKAARISVIEDEPLLEPGTTRAAQELDAIIGWTRAAIIGGAD 180

50 Query: 181 ATSNVRAGKIFNIPVSGTHAHLVQTYGDDYQAFKAYAEHTKDCVFLNDYDTLRGVGVN 240
 ATSNVRAGK F+IPVSGTHAHLVQ YG+DY AF AYA-THKDCVFLNDYDTLR+GVGF
 Sbjct: 181 ATSNVRAGKRFDIPVSGTHAHLVQAYGNDYAFMAKAYAEHTKDCVFLNDYDTLRGVGVN 240

55 Query: 241 AIRVAKEMGKINFLGVRLSDGLAYLSKKVRQQLDDAGFPNAKIYASNDLENTIIMLK 300
 AIRVAKEMG+KINFLGVRLSDGLAYLSK VRQQLDDAGF AKIYASNDLENTIIMLK
 Sbjct: 241 AIRVAKEMGDKINFLGVRLSDGLAYLSKTVRQQLDDAGFTPAKIYASNDLENTIIMLK 300

Query: 301 MQAKIDVWGVTGLITAYDQFALGAVYKIVSIETDAGSMRDTIKLSNNAEKVSTPGKQ 360
 MQAKIDVWGVTGLITAYDQFALGAVYKIVSIE + GSMRDTIKLSNNAEKVSTPGKQ
 60 Sbjct: 301 MQAKIDVWGVTGLITAYDQFALGAVYKIVSIETDAGSMRDTIKLSNNAEKVSTPGKQ 360

Query: 361 VWRITSRANGKSGDYITFAIDVQLDIEIMFHTPTTYTIINKIVRDFDAVPLLDVDFPKG 420
 VWRITSR KGSBGDYITF D+V +L EIMFHTPTTYI KIV+PDA+PLLDVDF KG
 Sbjct: 361 VWRITSRANGKSGDYITFDTINNELTEIMFHTPTTYTIINKIVRDFDAVPLLDVDFPKG 420

65

-1058-

Query: 421 KLVTQLPSLQEIQYGRKEFDQLWDEYKRVLPQDYFVDLARDVWQNMOLDIRIRKEAL 480
 +LVYQLP+L EI+ Y +KEFD+LWDEYKRVLPQDYFVDLARDVWQNM LID IRK+A
 Sbjct: 421 KLVTQLPFLAEIKAYAKKEFDKLWDEYKRVLPQDYFVDLARDVWQNMALIDIRKDAY 480

5 Query: 481 AKGR 484
 K B
 Sbjct: 481 GKSR 484

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 963

A DNA sequence (GBSx1021) was identified in *S.agalactiae* <SEQ ID 2939> which encodes the amino acid sequence <SEQ ID 2940>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2744 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GF:AAc74810 GB:AB000269 NAD synthetase, prefers NH3 over glutamine
 [Escherichia coli K12]

Identities = 173/274 (63%), Positives = 214/274 (77%), Gaps = 1/274 (0%)

Query: 1 MTLQDQIIKELGKVPINFGQIRRSVEFLKDYLLKHSEFLKTYVLGIGSQQDSTLAGRLA 60
 MTLQ QIIK LG KP IN +EIRRSV+FLK YL + F+K+ VLGI SQQDSTLAG+L
 Sbjct: 1 MTLQDQIIKALGAKPKINASEEIRRSVDFLKSGLYLTYPFIKSLVLGIGSQQDSTLAGLCLC 60

Query: 61 QLAVEELRADTG-ENYQFIATIRLPYGIQADESDAQKALDFIKEDIALTNINKEAVDQQRV 119
 Q+A+ ELR +TG E+ QFIA+RLPYG+QADE+D Q A+ FI+ED LT+NIK AV +
 Sbjct: 61 QVAINELELETONESLQFIATIRLPYGIQADESDAQKALDFIKEDIALTNINKEAVLASEQ 120

Query: 120 ALNPAAGVEITDFPNKNIKARQRMISQYAVAGQYAGAVIGTQHAARNTGFFTKRGDGGAD 179
 AL NG+E+DF +GN KAR+RM +QY++AG +G V+GTDHAAR ITGFFTK+GGG D
 Sbjct: 121 ALREAGIELSDFVRGNEKARERMKQYSGINAGTSGVVVGTGTHAARITGFFTKYGGDGT 180

Query: 180 LLEFLFRINKSQGKQLLAEIGADKALYKGIPTADLEENKPGIADIEALGVTYQETDAYLEG 239
 +EL+RLNK QGKQLLA L + LY+K FTADLE++P + DE+ALGVTY ID YLEG
 Sbjct: 181 INFLYRLNKRQSKQLLAALACPEHLKYGAFDTADLEDRPSLPEVALGVYTDNIDYLEG 240

Query: 240 KVVSDKSRGIENWYKQKHRLPITIFDDPWK 273
 K V + IENW+ K +HKR PIT+PDDPWK
 Sbjct: 241 KVVQGVARTIENWYLKTEHKRRPPIYVDDPWK 274

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2941> which encodes the amino acid sequence <SEQ ID 2942>. Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3482 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 213/274 (77%), Positives = 242/274 (87%), Gaps = 1/274 (0%)

-1059-

Query: 1 MTLDQQLIKELGVKPVINPSQIRRSVFLKDYLLKHSFLKTYVLGISGGQDSTLAGRLA 60
 MTLQ+II++LGVK I+P +EIR++V+FLK YL KESFLKTYVLGISGGQDSTLAG+LA
 Sbjct: 15 MTLQESTIRQLGVKASIDPQESIRKAVDFLKAYLRKHSFLKTYVLGISGGQDSTLAGRLA 74

5 Query: 61 QLAVERSLADTGRN-YQFIATRLFYGIQADSRDAQKALDFIKPDIALITINIKRAVNGQVR 119
 Q+A+ BLR + + YQFIATRLFYG+QADE DAQKAL FI PD LITINIK AVNGQV
 Sbjct: 75 QMAIARLRERASDAQYQFIATRLFYGVQADEADAQKALFIAPDQILITINIKRAVNGQVR 134

10 Query: 120 ALNAGVETIDPNKGNIKARQRMISQYAVAGQYAGAVIGTTHAENITGFFTFKFGDGGAD 179
 AL AAGVETIDPNKGNIKARQRMISQY+AGQ AGAVIGTTHAENITGFFTFKFGDGGAD
 Sbjct: 135 ALQAGVETISDPNKGNIKARQRMISQYAIAGQAGAVIGTTHAENITGFFTFKFGDGGAD 194

Query: 180 LLFLFRANKQGKQLIAELGADKALYEKIPADLEENKPGIADSRIALGVTOEIDAYLES 239
 +LFLFRANK QGK LL LGAD ALYEK+PTADLE+ KPG+ADE+ALGVTOY+ID YLES
 Sbjct: 195 LLFLFRANKQGKALLKVLGADAALEYEKVPTADLEDKPGIADSRIALGVTOEIDAYLES 254

15 Query: 240 KVSDEKSRGLIENWYKGGKRRHLPTITIDDFWK 273
 K++S ++ IE WY+GGKRRHLPTITIDDFWK
 Sbjct: 255 KLISKVAQATIEKRWKGGKRRHLPTITIDDFWK 286

20

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 964

A DNA sequence (GBSx1022) was identified in *S. agalactiae* <SEQ ID 2943> which encodes the amino acid sequence <SEQ ID 2944>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.2718 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:CA82960 GB:Z30315 aminopeptidase C [Streptococcus thermophilus]
 Identities = 363/444 (81%), Positives = 407/444 (90%)

Query: 1 MSKLTQTFPIKLFADYQANTKPSAIEANVTHNGLLESLETRQSEIENDYVFSIDLYKDEV 60
 M+ L+ FT+KLFADY+AN K+ AIENAVTHNGLLES+ETROSE+END+VFSIDLYKDEV
 40 Sbjct: 1 MTSLETDFTKLFADYENAKYGAIENAVTHNGLLESLETRQSEIENDYVFSIDLYKDEV 60

Query: 61 SNQKQSGRCWMPAALNTFRHLISDFKLENFELSQAHTFFWDKYKESNWFNQIATPQ 120
 SNQK SGRCWMPAALNTFRHLISDFKLE+FELESQAHTFFWDKYKESNWF+EQIATP+Q
 Sbjct: 61 SNQKASGRCWMPAALNTFRHLISDFKLESFELSQAHTFFWDKYKESNWFLEQIATPQ 120

45 Query: 121 ELSSRKVKFLLDVPOODGGQDMVVALFESKYGVVPKTVYPESVSSASRELNQYLNKILR 180
 E+ SRKVKFLLDV POODGGQDMV+LFSKYGVVPK+VYPSV+SS SRELNQYLNKILR
 Sbjct: 121 EIGSRKVKFLLDVPOODGGQDMVVALFESKYGVVPKTVYPESVSSASRELNQYLNKILR 180

50 Query: 181 ODAQILRELIAQAGDAGVQNKRELLQRIENFLNMILGLPQSGDFDAYRDKDNIHYQSDK 240
 ODAQILR+LIA GAD A VQ KKKK LQRIEN+LAM LGLP+ FDPAYRDKD++Y+S+K
 Sbjct: 181 ODAQILRDLIASGADQAQVQKKKEFLQRIENFLNMLGLPQSGDFDAYRDKDNIHYSEK 240

55 Query: 241 NITPKAFYQKYVNLDSYVSLINAPTVDKPYQSQSYTVEMLGNVVGSPAVKYLNLDMKRF 300
 ITP+AF++KYV L LSYDSV+INAPT DKPYQ+SYTVEMLGNVVG P+V+Y+NL M RF
 Sbjct: 241 GITPRAFFESKYVGLKLSYVSVINAPTADKPYGKSYTVEMLGNVVGSPAVRYINLPMERF 300

Query: 301 KELATAQMKSGETVWFGSDVQGVSNRQKGIATNTTYDFPMSMDIKLSQDKAGRLDYSES 360
 KELATAQMK+GE+VWFGSDVQGV+S+RQKGIAT YDF +SMDI +QDKAGRLDYSES
 60 Sbjct: 301 KELATAQMKAGESVWFGSDVQGVSDRQKGIATNTTYDFPMSMDIKLSQDKAGRLDYSES 360

Query: 361 MTHAVVLITGVLDSESGQLNKVKNVSGESKVGKDGYPVASDAMDEYTYQIVVRKELLTK 420

-1060-

MTHAMVLTVGVLDL G+P+KWK+KNSWG+KVG+ GYFVADAMMDREYTYQIVVRK+ LT
 Sbjct: 361 MTHAMVLTVGVLDLADGKPIKWKIENSWGDKVGQKGYFVADAMMDREYTYQIVVRK+DLTA 420

Query: 421 EELEAYNAEPITLAPNDPMGALAN 444
 EEL AY A+P LAPNDPMG+LA+
 Sbjct: 421 EELEAYEADPQVLAPNDPMGSLAS 444

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2945> which encodes the amino acid sequence <SEQ ID 2946>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3002 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 369/443 (83%), Positives = 407/443 (91%)

Query: 1 NSKLTQTFTDKLPADYQANKFKSAIENAVTINGLKLSLETQSEIENDYVPSIDLTQDEV 60
 NS LT+TFT++LFA Y+AN KFSAIENAVTINGLKLSLETQSE++ND+VPSIDLTQD+V
 Sbjct: 1 NSALTETFTQELFAHYEANKFSAIENAVTINGLKLSLETQSEVNDPVSIDLTQDKV 60

Query: 61 SNQKSGRCMMPAALANTFRHKLISDFKLENFELSAHTFFWDKRYEKNWFMEQ+IAT+Q 120
 SNQK SGRCMMPAALANTFRHKLI+FKLENFELSAHTFFWDKRYEKNWFMEQ+IAT+Q
 Sbjct: 61 SNQKASGRCMMPAALANTFRHKLITEFKLENFELSAHTFFWDKRYEKNWFMEQVIATDQ 120

Query: 121 ELSGRKVKFLLDVPQDQGCQCMNVVALFEKYGVPKTYVPESVSSASRELAKVNLKLR 180
 EL+GRKVKFLLDVPQDQGCQCMNV+LFEKYGVPK+VYPS+SSS SRELAKVNLKLR
 Sbjct: 121 ELTSRKVKFLLDVPQDQGCQCMNVSLFEKYGVPKSVYPSISSNSRELAKVNLKLR 180

Query: 181 QDAQILRELIAQGADGATVQNKKEELQRIENFLAMHGLPQSDPAYRDKDNHYQSDK 240
 QDAQILR+LIA GA V++K ELLQEIENFLAM LGLPP+ FDFAYRDKD+HY +K
 Sbjct: 181 QDAQILDLIASGAKADQVEERKAELLQEIENFLAMTLGLPPRHFDFAYRDKDNHYHVEK 240

Query: 241 NITPKAFYQKYVNLDSLSDYVSIINAPTVDKPYGQSYTVEMLGNVGGPAVKYLNLMKRF 300
 +TP+AFY K+V L LSDYVS+INAPT DKPYG+SYTVEMLGNVVG V+YLNLMKRF
 Sbjct: 241 GLTPQAFYDKFVGLKLSYVSVINAPTADKPYGKSYTVEMLGNVVGSEVRYLNLMKRF 300

Query: 301 KELAIQMKAGSETVWPGSDVGQVSNRQKGLIATTTDPHNSMDI KLSQDKAGRLDYSES 360
 KELAI QM++GE+VWPGSDVGQV+RQKGLIAT TYDF +SMDI LQSDKAGRLDYSES
 Sbjct: 301 KELAIQMKAGSETVWPGSDVGQVSDRQKGLIATNTYDFEASMDINLSQDKAGRLDYSES 360

Query: 361 MTHAMVLTVGVLDLSEGGPLKWKIENSWGKVGKGYFVADAMMDREYTYQIVVRK+LT 420
 MTHAMVLTVGVLDLSE+G+PLWKIENSWGKVGK GYFVADAMMDREYTYQIVVRK+LT
 Sbjct: 361 MTHAMVLTVGVLDLSTGKPLWKIENSWGKVGKGYFVADAMMDREYTYQIVVRK+DLTA 420

Query: 421 EELEAYNAEPITLAPNDPMGALA 443
 +EL AY EP LAPNDPMGALA
 Sbjct: 421 EELEAYEKEPQVLAPNDPMGALA 443

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 965

A DNA sequence (GBSx1024) was identified in *S.agalactiae* <SEQ ID 2947> which encodes the amino acid sequence <SEQ ID 2948>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have a cleavable N-term signal seq.

-1061-

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9533> which encodes amino acid sequence <SEQ ID 9534> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF17262 GB:AF210752 penicillin-binding protein 1A
 [Streptococcus pneumoniae]
 Identities = 412/725 (56%), Positives = 544/725 (74%), Gaps = 14/725 (1%)

Query: 4 IKKESVIKLLKYPAGIIMQPIIAIVIGQLFYVVSRSFKLTDQALKSVNSIIVDGN 63
 + K ++++L+KY + +I AI+GG +F YVBS++P L+ L + SS +YD N
 Sbjct: 1 MNKPTILRLIKYLSISPLSLVIAIVIGGGVFYVSKAPLSBSKLVATTSSKIYNNK 60

Query: 64 KLIADLGSEKRSVSADSIPLNLVNAITSIEDKRFFHNGVDIYRILGAAMNVLSSNTQ 123
 +LIADLGSE+R + A+ IP +LV AI SIED RFF HNG+D RILGA NL SS +YD N
 Sbjct: 61 QLIADLGSEKRSVSADSIPLNLVNAITSIEDKRFFHNGVDIYRILGAAMNVLSSNTQ 120

Query: 124 GGSITLQQLIKLAYPSTNKSQDTLKKRSQEVWLAQMERKITRBEILTFYINRYHNGN 183
 GGSITL QQLIKL YPST+ SDQT+ RK+QE WLA+Q+E+K TK+ELIT+YINKYIM HKN
 Sbjct: 121 GGSITLQQLIKLAYPSTNKSQDTLKKRSQEVWLAQMERKITRBEILTFYINRYHNGN 180

Query: 184 YGMRTTAKSYFGKDLKELSLAQALLAGIPQAPTQIDPYNFSPESAGTRNTVLQKYQDK 243
 YGM+T A++Y+GKDL LS+ QALLAG+PQAP QIDPY +PE+AQ RRN VL +
 Sbjct: 181 YGMRTTAKSYFGKDLKELSLAQALLAGIPQAPTQIDPYNFSPESAGTRNTVLQKYQDK 240

Query: 244 NISKKEVDQAVATPVTDLKELKQKSTYPKMDNYLKQVISVSKQKTGKIDFTAGLYVT 303
 IS ++Y++AV TP+TDGL+ LK S YP YMDNYLK+VI++V++TG ++T G+ VYT
 Sbjct: 241 YISBQTEKAVNTPTITDGLQSLKASNYPAYMDNYLKEVINQVEETGYNLLTGMVYT 300

Query: 304 NIMTDQAKQLYDINSDTYIAYPNMELQIASTINDATNGKVIAQLGQRHQNENISRGTNQ 363
 N++ +AQK L+DIYN+D Y+AYP++ELQ+ASTI+D +NGKVIAQLG RHQ+ N+SPG NQ
 Sbjct: 301 NVDQERAKHLMDIYNDEYVAYPDDELQVASTIVDSNGKVIAQLGARHQSNNVSGINQ 360

Query: 364 SVLTNRDNGSTMKPIISAYADAIDGVNMTQSLNDSVYVPGTSTQLYDNDQRYMGRS 423
 +V T+RDNGSTMKPI+ YAPA++ GUY+ST ++D Y +PGT T +Y+NDR Y G ++
 Sbjct: 361 AVEINRDNGSTMKPIITDYAPALEYGVYDSTATIVHDEPNYKGTDPVYNDRGYFGHIT 420

Query: 424 MQTAAIQSRNVPAYRAALEAGLDEAKSFLKLGITTYPEMYNSNAISSNSSDRAKYGASS 483
 +Q A+QSRNVPAY L GL+ AK+EL LGI YP ++YNSNAISSN +SD KYGASS
 Sbjct: 421 LQVALQSRNVPAYETLANKVGLNRKATFLNGLGIDYPSLHYSNAISSNTTSDKKYGASS 480

Query: 484 EKGAAAYAPANGGTYYKPOYVKNIEPDSQDNTDYAASGRSAMKEKTTAYMMIDMLKTLVLT 543
 EKGAAAY+APANGGTYYK Y++K+ PSDG+ ++ G+RAMKEKTTAYMMIDMLKTLVLT
 Sbjct: 481 EKGAAAYAPANGGTYYKMYTHKVFSDGSEKPSVGTGRAMKETTA YMMIDMLKTLVLT 540

Query: 544 PGSTGKAAIPGVAGKRGTSNYTSYDRIKAKATGIIYNSAVGTMAPDKNFVGYTSKYTM 603
 +G G A +P + QAGKTGTSNYT++E+ K Y G +APDE FVGYT KY M
 Sbjct: 541 YGIGRGAYLFWLPQAGKTGTSNYTDEEIKR-----YIKNIGYVADPEMFVGYTKYAM 593

Query: 604 AINTGYENRLTPLYGSDLIATVYRAMSYLTQGYSA-DWIMPRGLYRSGSYLYINGTT 662
 A+WICY NRLTPL G L +A +VYR+M+YL+ G + DW +PGLYR+G +++ NG
 Sbjct: 594 AVNTGYENRLTPLYGSDLTVAAKVYRSMYLYLSESGNPELWNIHGLYRNGEPFVKNGAR 653

Query: 663 TTGTYSSEVYKNIYQNSQSSQSSSSTSEKQEDKEDYANDANSSSPQVETPNNGKATTP 722
 +T +SS + S +SS SSS +S+ + N++ ++P T + TTP
 Sbjct: 654 ST--KSSPAQC--PPSTESSSSSSSSTSGSSTTTPSTNASTTTNNNNQQSN--TTTP 707

Query: 723 NNSNQ 727
 + NQ
 Sbjct: 708 DQCNQ 712

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2949> which encodes the amino acid sequence <SEQ ID 2950>. Analysis of this protein sequence reveals the following:

```

Possible site: 41
>> Seems to have an uncleavable N-term signal seq
5  INTEGRAL    Likelihood =-13.96    Transmembrane    19 - 35 ( 9 - 43)

----- Final Results -----
          bacterial membrane --- Certainty=0.6583 (Affirmative) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
10         bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:CA88918 GB:249095 penicillin-binding protein 1a [Streptococcus pneumoniae]
15  Identities = 422/712 (59%), Positives = 536/712 (75%), Gaps = 8/712 (1%)

Query: 4   IKNPKILKWLKYLKLSLILVLIIGGLLFYIISAPKLSAQKSTNSSLIYVDNN 63
      + P IL+ +KY+ + LSL+I I+GG +F +Y+S AP LSE++L +T SS +YD N
Sbjct: 1   MNKPTILRLIKYLSISPLSLVIAIAVLGGVFFYYSKAPLSBSKLVATTSSKIYDNN 60

Query: 64  NLIALDLSEKRNVTASIPINLVNAITSIEDKRFPHRGVDLYRIFGAAPHNLTSQTQ 123
      LIALDLSE+R N A+ IP +LV AI SIED RFF+HRG+D RI GA NL S + Q
Sbjct: 61  QLIADLSEKRRVNAQANDITDVLKAIIVSIEDHRFFDHRGIDITRIILGAFLENLQSNLQ 120

Query: 124  GGSTLDQQLIKLAYFSTNESDQTLKKAQEBVWLALQNERKYTKQELIFYINKYVMNGN 183
      GGSTL QQLIKL YFST+ SDQT+ KKAQE WLA+Q+E-K TKQELIT+YINKYVM NGN
Sbjct: 121  GGSTLYQLIKLITYFSTSTSDQTSIRKAQEAWLAIQLEKATKQELITYINKYVMNGN 180

Query: 184  YGMITAAKSYGKDLKDLQALLAGIQAPSQYDPYLIHPEAAQRNRNVLLQMYMEK 243
      YGM TAA++Y+GKDL +LS QLALLAG+QPAF+QYDPY HPEAAQ+RRN+VL +M +
Sbjct: 181  YGMQTAQNYGKDLMLNLSLPQLALLAGMPQAPWQYDPYSHPEAAQRNRNLVSEKNGQ 240

Query: 244  HLTYGAEYETALTPVAEGLQSLQQRSTYPKYMDNYLKQVIEVKKETNKDIFTAGLKVYT 303
      +++ +YE A+ TP+ +GLQLS+ S YP YMDNYLK+VI +V++ET ++ T G+ VYT
Sbjct: 241  YISAEQYEKAVNTPTITDGLQSLKASNYPAYMDNYLKEVINQVEETGYNLITG+DVYT 300

Query: 304  NIIPDAQQTLYNIYHSGDVIYYPDQDFQVASTIVDVNTHVLAQLGGRNDENVSPTNQ 363
      N+ +AQ+ L++IY+S YV YPD D QVAST+VDV+NG VIAQLG R+Q NVSPGTNQ
Sbjct: 301  NVDQEAQGLHMDIYNSDQVSYFEDDLQVASTVVDVSNKVIQALGARQASNVSPGTNQ 360

Query: 364  AVLITRDWGSMTKPIATAPAIRSGVYTTAQSTINDSVYWGTTITQLFNMLDLYNGMT 423
      AV T+RDWGS+MKPIT YAP+A+L GLD A++FL+ LGI+Y MHY+NAISEN +S+KYGASG
Sbjct: 361  AVETNRDWGSSMKPITDYAPALEYGVYDSTASIVEDVPYIPGTDTPLYNMDHYVFGNIT 420

Query: 424  IQAAILSRNVPAVRALEAGLYARSFLSSLGINYPMHYNSAISNNSSDDKYKQASS 483
      IQ A+ SRNV AV L GLD A++FL+ LGI+Y MHY+NAISEN +S+KYGASG
Sbjct: 421  IQYALQSRNVTAVENTLKVGLDRKFTLGLGIDYPSMHTANAISNTTSSNKKYKQASS 480

Query: 484  EKMAAAYAAPFANGGIYHKPRYNKVEFSDGTSKTFDEKOKRAMKETAYMMTDLKITLT 543
      EKMAAAYAAPFANGGIYHKP +Y+NK+ FSDG+ K F + G RAMKETAYMMT+M+KITVLT
Sbjct: 481  EKMAAAYAAPFANGGIYHKPMYINKIVFSDGSEKFPDAGTRAMKETAYMMTMEKIVLT 540

Query: 544  YGTGTAAAIQVGAQKGTGSNTYDEELAKIGEKGLYPDVGTILAPDENFVGPTKRYAM 603
      YGTG A +P + QAGKGTGSNTYDEE+ K Y G +APDE FVG+T+YAM
Sbjct: 541  YGTGRGAYLPLVPAQKGTGSNTYDEEIEK -----YIKGTGVVAPDEMFVGPTKRYAM 593

Query: 604  AVMTGYNRLTPIPVYSSSLIASDVSRRMTYLT-NGYSEDWTPMNGLYRSGGFLYLSOTY 662
      AVMTGY NRLTP+ G A+ VYRSM+TYL+ + DWTPM+GLYR+G F++ +G
Sbjct: 594  AVMTGYNRLTPIIGDGLVAGKRYRSMITLYLSEDDQFGDWTPMGLYHNGEYFFFGAR 653

Query: 663  ASNTDYNSVNNLYSNMTTASSQTSDDTSNNSNTNSNTNNGNSHPST 714
      ++ + + S++++ SS + S+ T+ S + S + S +T+NN +T
Sbjct: 654  STWSSPAPQPPSTSSSSSSSDSTSQSNSTTPSTNNTTINPNNTQQSNPT 705

```

An alignment of the GAS and GBS proteins is shown below.

-1063-

Identities = 521/729 (71%), Positives = 621/729 (84%), Gaps = 10/729 (1%)

Query: 1 NITIKKESVILKLYAFGIIMGFIIAIVIGILLFAYTVSRPKLTDQALKSVNSSLVYD 60
 +ITIK ++K LKY I+ IIL I+IGILLF+Y+S+PKL++ LKS NSSLVYD
 5 Sbjct: 1 VITIKGKILKMLKYVLGAILSLIILVIIIGILLFTFYISSAPKLEAQKSTNSSLVYD 60

Query: 61 GNNLIADLGSEKREVSADSIPLINVAITSIEDKRRFFKRGVDIYRILGAAMHNLVSS 120
 GNN LIADLGSEKRE+V+ADSIPLINVAITSIEDKRRFF HRGVD+YRI GAA+HNL S
 10 Sbjct: 61 GNNLIADLGSEKRENVTADSIPLINVAITSIEDKRRFFHNRGVDLYRIQGA+HNLTSQ 120

Query: 121 NTQGGSTLDQQLIKLAYPSTNKSQTLKRRKQGVWALQMERKRYTKRILTFYINKVYMG 180
 TQGGSTLDQQLIKLAYPSTN+SDQTLKRRK+QEWALQMERKRYTKRILTFYINKVYMG
 20 Sbjct: 121 TQGGSTLDQQLIKLAYPSTNKSQTLKRRKQGVWALQMERKRYTKRILTFYINKVYMG 180

Query: 181 NGNYGMRTTAKSYPGKDLKELSLAQALLAGIIPQAPTQYDPYKNPESAQTRRNTVLQPMY 240
 NGNYGM T AKSY+GKDLK+LS AQALLAGIIPQAP+QYDFY +FE+AQ RNN VLQPMY
 25 Sbjct: 181 NGNYGMLTAAKSYGKDLKELSLAQALLAGIIPQAPSQDTPYLHPEAQRRNVVLQPMY 240

Query: 241 QDQNISSKKEYDQAVATPVTDGLKELQKSTYPKYMNYLKQVISEVKQKTGKIDPTAGLK 300
 +K++K EY+ A+ATPV +GL+ L+Q+STYPKYMNYLKQVI EVK++T KIDPTAGLK
 30 Sbjct: 241 MEGLTKAEYETAIAIPVABGLQSLQQRSTYPKYMNYLKQVISEVKRETNDIPTAGLK 300

Query: 301 VYTNIINDAQQLYDINSYTIAYPNKLQIASTIMDATNGKVIQAQGRGHONENISFG 360
 VYTNI DAQ+ LY+I+Y+S Y+ YP+ +Q+ASTI+D TNG VTAQGR+Q+HEH+SFG
 35 Sbjct: 301 VYTNIIPDAQQLYNTYIHSQDVIYVYDQDPQVASTIVDVTNGHVIQAQGRQDEHVSFG 360

Query: 361 TNQSVLTDPRDWGSTMKPIAYAPAI+SGVY ST QS NDSVYVWEGT+TQI++ND +G 420
 TNQ+VLTDPDWGSTMKPI+AYAPAI+SGVY ST QS NDSVYVWEGT+TQI++ND +G
 40 Sbjct: 361 TNQAVLTDPRDWGSTMKPIAYAPAI+SGVYVTSQAQSTNDSVYVWEGT+TQI++ND+KNG 420

Query: 421 WMSMQTAIQGSRNVPVAVRALEAAGLDEAKSFLKLGIIYPPMNYNSNAISSNNSSDKKYG 480
 WM+AQ AI SRNVPVAVRALEAAGLD A+SFL LGI YPM+YNSNAISSNNSSDK KYG
 45 Sbjct: 421 WMTIQAAIMLSRNVPVAVRALEAAGLDVARSFLSGIINTPEMHSNAISSNNSSDKKYG 480

Query: 481 ASSEKGAARYAFANGGTYKPYVNMKIEFSDGINTYAAAGSRAMKETTAYMTIMLMT 540
 ASSEKGAARY+AFANGG Y+KP+YVMK+EFSDGT+ T+ G RAMKETTAYMTIMLMT
 50 Sbjct: 481 ASSEKGAARYAFANGGIYHKPRYVNMKIEFSDGTISKTDFEKGKRAMKETTAYMTIMLMT 540

Query: 541 VLTFGTGTAAALPGVAQAGKTQTSNTYTEDELAKEIATGTIYNSAVGTMAPDENFVGYSK 600
 VLT+GTGT AAIPGVAQAGKTQTSNTY++ELAKI G+Y VGT+APDENFVG+T +
 55 Sbjct: 541 VLTGTGTAAALPGVAQAGKTQTSNTYTEDELAKEIYKGLYDPYVGTMAPDENFVGFTKR 600

Query: 601 YTMALWGYKRLTPLYGSSOLDIATEVYRAMMSYLTGGYSALWTPMEGLYRSGSYLYING 660
 Y MA+WTGYKRLTP+YGS L+IA++VYR+MM+YLT GYS DWTPM GLYRSG +LY++G
 60 Sbjct: 601 YAMAWWGYKRLTPVYGSSELIASDVYRMMTYLTNGYSEDWTPMNGLYRSGGFLYLSG 660

Query: 661 TTTTGT-YSSSYVKNYQNSQGSQSSSTSEKQKEDONTANDANSSSQVETPNNGNA 719
 T + T Y++SVY N+Y N ++++ SS+ +D +++ND ++S+ T NGG+
 65 Sbjct: 661 TYASNTDYNSVYNNLYN-----NITTASSQTTSDTSSSNNTYNSNT---NTDNNSH 711

Query: 720 TTPNNNSQT 728
 + ++ T
 Sbjct: 712 PSTDDKKT 720

A related GBS gene <SEQ ID 8695> and protein <SEQ ID 8696> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 MoG: Discrim Score: 6.55
 GVH: Signal Score (-7.5): -1.98
 Possible site: 36
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 0 value: 4.03 threshold: 0.0
 PERIPHERAL Likelihood = 4.03 201
 modified ALOM score: -1.31

*** Reasoning Step: 3

-1065-

2157 2187 2214 2244 2274 2304 2334 2364
 KNRLLTPLYGSGDIDATEVFRMKAITL-GGYSADWTMEPEGLRYGTYLIMGTITITGYTSVYKNIQNSQGSQSSSS
 ||||| : : :|||:|:|:| : || :|||:|: : : : : : : : : : :||| :
 SNRLITPLYGDLTVAARVSNMTLLTSSGSEPELWNIPEGLRYNGRVPFKNGARST--WSPAPQQ--PPSTSSSSSSSS
 610 620 630 640 650 660 670
 2394 2424 2454 2484 2514 2544 2574 2604
 TSSGKQKEDITANDANSSSGVFEYRNGNANITPNSNQTIVGTGHGNGNSNNNTVNGR-KTGYIIQFFNI-LELLI
 ||||| : : :|||:|:|:| : || :|||:|: : : : : : : : : : :||| :
 SETSGQSNSTPTSTNSITNTINPNTQGS--NTTPEQQNQMPQAPQ
 690 700 730

SEQ ID 8696 (GBS146) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 23 (lane 4; MW 82kDa), in Figure 168 (lane 11-13; MW 96.5kDa) and in Figure 238 (lane 8; MW 96.5kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 2; MW 107kDa).

Purified Thio-GBS146-His is shown in Figure 244, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

20 Example 966

A DNA sequence (GBSx1025) was identified in *S.galactiae* <SEQ ID 2951> which encodes the amino acid sequence <SEO ID 2952>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3647 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA26957 GB:M90528 ORF [Streptococcus oralis]
Identities = 143/196 (72%), Positives = 165/196 (83%), Gaps = 1/196 (0%)

35	Query: 1	MVNYPHQLIRKTTVTYSKKKKIKDIPNRGMSFEADINATDITDYLSHGLAVIHKKPTPQIV 60
		MVNYPH++ ++ + + K + PANRGRMS DIATDITDYLSHGLAVIHKKPT+QIV
	Subject: 1	MVNYPHCLISGQAPPCSTK-NFANRGRMSFEKMINATDITDYLSHGLAVIHKKPTPQIV 59
	Query: 61	KVDYPKRSRAKIVYAPFQASTDYSGVGYGYIIDFEAKETRQKTAAMPKNFPAHQIEHM 120
		+VDY+PKRSRAKIVYAPFQASTDYSGVY GYIIDFEAKETRQK+A+PMKNF HQI+HM
40	Subject: 60	KVDYPKRSRAKIVYAPNQASTDYSGVGYIIDFEAKETRQKHAHPMKNFPHHQIHEQM 119
	Query: 121	ANVLQKGQICFVLPHLSITKETYLLPANELISFYQIDKGNKEMPIDYIRKNGFFVKESAF 180
		VL Q+GICFVLPHLF++ +ETYLLEA +LI FY DKG KEM+P YIR+NG+ ++ AF
45	Subject: 120	EQVLQAGQICFVLPHFASQCEITYLLPAVDLIRFYHQDKQKQSMPLGYIRENGYRIELGAF 179
	Query: 181	POVFDIIEEIKLGG 196
		PQ+VFDIIE+K LGG
	Subject: 180	POIFVDIIEEELGG 195

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2953> which encodes the amino acid sequence <SEQ ID 2954>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
```

-1066-

bacterial cytoplasm --- Certainty=0.5030 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 166/199 (83%), Positives = 177/199 (88%)

Query: 1 NVNYPHQLIRKTTVTKSKKKKIDFANRGMFSAADNATNDYLLSHELAVIHKKKTPVQIV 60
 10 Sbjct: 1 NVNYPH LIR+ + K+ K+DFANRGMFSAADNATNDYLLS ++AVIHKKKTPVQIV 60
 Query: 61 KVDYFKRSRAKIVRAYFRQASTTDYSGVYKGYIDFEAKETRGKIDAMPKMFHQAIEHM 120
 Sbjct: 61 KVDYFKRSRAKIVRAYFRQASTTDY GVTG+Y+DFAKETRGKIDAMPKMFH QAIEHM 120
 15 Query: 121 ANVLQCKGICFVLLHFSTLAKETYLFPANELISFTQIDKONKSMFIDYIRKNGFFVKESAP 180
 Sbjct: 121 ACVLHQRGICFVLLHFSTLAKETYL LFA LISFTQID G+KSMFIDYIRKNGF V AF 180
 20 Query: 181 PQVPYLLIIEKLLGGDIN 199
 Sbjct: 181 PQVPYLLIIEKLLGGDIN 199

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 25 vaccines or diagnostics.

Example 967

A DNA sequence (GBSx1026) was identified in *S. agalactiae* <SEQ ID 2955> which encodes the amino acid sequence <SEQ ID 2956>. Analysis of this protein sequence reveals the following:

Possible site: 61
 30 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3227 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CB14136 GB:Z99115 similar to hypothetical proteins from B. subtilis [Bacillus subtilis]
 40 Identities = 74/174 (42%), Positives = 97/174 (55%), Gaps = 6/174 (3%)
 Query: 5 ILVTGYKNFELGIPQKQDPRTIILKKAIDKFRPLENGADWFIEMGLGFEYNALVAL 64
 + +RGYK FELGIP+ D + IKKAI FL+ G +W + G LG E NA E A
 Sbjct: 4 LAITGYKFFELGIPKQDQKALYIYIKAIKNRLIAPLDEHLEWLLISGQLGVLAQAEAY 63
 45 Query: 65 DLQKEY-DPQIATITFTFENHGNWANKAKL-ALFKQVDF-VKYTFPSYENFGQPKQYN 121
 DLQ+RY D ++A I F +NW E NK + A+ Q D+ T YR+P QFNQ N
 Sbjct: 64 DLQREYFDLKVAVITFPFYKERNWKNPKGQYEAVALQADYEAULTHRPYESPLQFKQKN 123
 50 Query: 122 HFLINNTQAYLFYDSKRNKTNLKFLLRMEKK---HAYDISPITFDRLNIYIE 172
 F I+ + G L YD E E + K++L EK+ + Y I F+T D L EE
 Sbjct: 124 QFFIDKSDGLLLLYDPEKSGSKYMLGTAKRRSQDGYFTYITFMDRLRYVES 177

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2957> which encodes the amino acid
 55 sequence <SEQ ID 2958>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----

-1067-

bacterial cytoplasm --- Certainty=0.3041(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 5 An alignment of the GAS and GBS proteins is shown below.

Identities = 102/167 (61%), Positives = 127/167 (75%)

Query: 3 STILPTCYGNFELGIFQDKDRITITIKKAIKDKPRFLEKADWFIEMGNLGFYEWALNV 62
 + IL+TGY++FE+GIF KDFR++ITK+AI KD +LENG DWFI QNLGFR WALEV
 10 Sbjct: 2 TAILITGVRSEFEGIFDHDKDFRVSIIKQAIRKDLIGYLENGVDWFIPTGNLGFQWALEV 61

Query: 63 ALDLQKSYDFCIATITPFENNHCQMMREANKAKALFKQVDFVKYITPFSEYENFGQFKQYNH 122
 A +L++EY QIATIF FE HQ WNE NK L+ P+ VDPVKY FP+YE P QF QY
 15 Sbjct: 62 ANELKEEYPLCIATITPLFETHGRMNEKNKEVLSQFRAVDVFKYIFYENYBQPTQFSQYQ 121

Query: 123 FLINNTQCAVLFYDSENETNLKFLLEMMSEKKEAYDISPLTFDRINET 169
 FL+ T+GAY+FYD+ENETNLK+ L+ + Y+ LTFDRIN++
 Sbjct: 122 FLLEKTEGAYVFYDTENETNLKYFLKAKAMPHYQLLLLTFRINAM 168

- 20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 968

A DNA sequence (GBSx1027) was identified in *S.agalactiae* <SEQ ID 2959> which encodes the amino acid sequence <SEQ ID 2960>. Analysis of this protein sequence reveals the following:

25 Possible site: 23
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.5188(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 969

A DNA sequence (GBSx1028) was identified in *S.agalactiae* <SEQ ID 2961> which encodes the amino acid sequence <SEQ ID 2962>. This protein is predicted to be cell division protein DivIVA. Analysis of this protein sequence reveals the following:

40 Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2736(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 50 A related GBS nucleic acid sequence <SEQ ID 9535> which encodes amino acid sequence <SEQ ID 9536> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1068-

>GP:CAB14135 GB:Z99115 ypeB [Bacillus subtilis]
 Identities = 46/102 (45%), Positives = 69/102 (67%), Gaps = 14/102 (13%)

5 Query: 14 SPKDI FEQDFKVMRGYDKKEVDVFLDDVIKDYENYLEQIEKLMQENRRLLQALDKKESE 73
 S K+I E++FK -RGY ++VD FLD +IKDYE + ++IE+LQ EN +L++ L+ E
 Sbjct: 9 SAKSILEKEFKTKGVRGKYQKQVDKFLDMIIKDYETPHQIEELQENLQLKQLE----E 64

Query: 74 ASNVVNSGTAMYNQKPIAQSATNFDILKRISRLKEKVFGRQI 115
 AS ++P+ + TNFDILKR+S LEK VPG ++
 10 Sbjct: 65 AS-----KKQPVQSNITNFDILKRISRLKEKHVFGSKL 96

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2963> which encodes the amino acid sequence <SEQ ID 2964>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4466 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 20 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 71/112 (63%), Positives = 85/112 (75%), Gaps = 6/112 (5%)

25 Query: 8 MASIIYSPKDI FEQDFKVMRGYDKKEVDVFLDDVIKDYENYLEQIEKLMQENRRLLQAL 67
 M SIYSPKDI FEQ+FK SMRG+DKKEVD FLD+VIKDYEN+ QLE L+ EN +AL
 Sbjct: 1 MTSIIYSPKDI FEQDFKTSNRGFDKKEVDFLDNVIKDYENFNAQIEALKAEN----EAL 56

Query: 68 DKKESEASNVNSGTAMYNQKPIAQSATNFDILKRISRLKEKVFGRQIRE 117
 K + +A N ++ +P +AQSATNFDILKRIS+LEKVFGR+QI E
 30 Sbjct: 57 KKAKFQARNTVTSATVQQPVQFQTRVAQSATNFDILKRISRLKEKVFGRQIIE 108

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 970

A DNA sequence (GBSx1029) was identified in *S.agalactiae* <SEQ ID 2965> which encodes the amino acid sequence <SEQ ID 2966>. Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence (or aa 1-19)

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0655 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 40 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14134 GB:Z99115 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 204/382 (53%), Positives = 274/382 (71%), Gaps = 3/382 (0%)

50 Query: 3 ESFKLIATAAGLEAIVGRIRIENIGIDCCQVNGRVRPHGDIKTIITKTNINIRAAADRIKII 62
 ++ LIIATA G+EA+V +E+R+LG +C+V+NG+V P GD I NIKLR ADRIK+
 Sbjct: 2 KKCTLIATAFPGIIEAVVAKEVRDLGYECKVDNGKVIPEGDALAICRANLWLTADRIKIVQ 61

Query: 63 VGEFFPAPTFEELPGQVYGLDWENYLPQAKPFIKAKCKVCKLHNHPSVQAISKKAVAKK 122
 V FA TP+ELF+ ++W ++P KFP+ K VKS L + P Q I KKA+ +K
 55 Sbjct: 62 VASFKATTFDELPEKTKAINWRSFIPENKFPVI-GKSVKSTLASVDCQRIKKAIVK 120

Query: 123 LQKVMIRFEGVPLQENGAIEKIEVSLKDKATVMITDTGSSLPKRGVRAEKGGAPIKENNI 182
 L K+ ++E GAE+K+E+S+LKD+A + +D+G+ L KGRYR ++GGAPIKE +

-1069-

Sbjct: 121 L-KLQSGKANDWIEBTGAEYKVBISLLKQALITLDSSTGLHKG YRVDQGGAPIKETL 179

Query: 183 AAIITQLSNWFPDKPLIDPTCGSGTFCIEAAMIGGNIAFGPNRDFAPZAMFWVDSQSQVK 242
AAA++QL+NW PD+P +DP CGSGT IEAA+IG NIAPGPNRDP E W W+ + K

5 Sbjct: 180 AALVQLINWTFDPFFVDFGSGTIAEALIGGNIAFGPNRDFVSDMEWIKGLWANK 239

Query: 243 VRDEASSKANYDIDLDSGFDLDGRWVEIARNAEAGLGDVTKLQMRQLDLATKTKNG 302
R E E KANYD L I D+D RMV+IA++NAEAGLGD+I+ KQM++D T+ G

10 Sbjct: 240 ARLVVEEKANYDQPLTIFASDIDHRWVQIAKENAEAGLGDLTQFKQMQVKDFTNMLERQ 299

Query: 303 VIISNPPYGERLLDDKAVDILYNEMQQTFAPLKTNWKFILTSDEGFEKKYGSQADKKRKL 362
VI+ NPPYGERLL + KAV+ +Y IEMQ ? FL TWB ++LTS+E FB+ YG +A KKKKL

15 Sbjct: 300 VIVGNPPYGERLGEKGAIVEQYKEMGQAFSPLOTWSVYMLTSNENFEAYGRKATKKRKL 359

Query: 363 YNGTLLKVDLYQYGERVRRQVK 384
+HG +K D YQY+ +VR Q R K

20 Sbjct: 360 FNGFIKIDYIYQW-SKVRQKR 380

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2967> which encodes the amino acid sequence <SEQ ID 2968>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0324 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 317/383 (82%), Positives = 354/383 (91%)

Query: 1 MKSFKLIATAAGLEAIVGRIRRIHTGDCQVNGVRVPHGDKITITINLMLRAADRIK 60
MKE+P+L+ATAAAGLEA+VG+E+R LG DCQVNG+V F CD++ I+TNLMLRAADRIK

35 Sbjct: 1 MKITFRLVATAAGLEAVVGKEVRALGDCQVNGKVFYFSDVBAIVKTNLMLRAADRIK 60

Query: 61 IIVGQFPARTFEELFGQVYGLDWEHYPLGAKFPIAKAKCVKSLHNEPSVQATSKKAVA 120
IIVG+PFA TFEELFGQV+ LDWEHYPLGAKFPI+KAKCVKSLHNEPSVQAT+KKAV

40 Sbjct: 61 IIVGQFPARTFEELFGQVFDLWENYPLGAKFPIAKAKCVKSLHNEPSVQATKKAVV 120

Query: 121 KKLQKVFHRPBGVPLQENGAEPKIEVSLKDWATVMIDTGSLLPKRGYRAEKGGAPKE 180
KKLQK FHRPBGVPLQF G+ F IEVSLKD+AT+MIDTGSLLPKRGYR +RGGAPKE

45 Sbjct: 121 KKLQKVFHRPBGVPLQFVGSTFWIEVSLKDWATVMIDTGSLLPKRGYRVQGGAPKE 180

Query: 181 NMAAAIITQLSNWFPDKPLIDPTCGSGTFCIEAAMIGGNIAFGPNRDFAPZAMFWVDSQSQV 240
NMAAAI+ LSNWFPDKPL+DPTCGSGTFCIEAAMIGGNIAFGPNRDFAPZ+V WVD+ V

50 Sbjct: 181 NMAAAIILSNWFPDKPLVDPTCGSGTFCIEAAMIGGNIAFGPNRDFAPZAMFWVDSQSQV 240

Query: 241 QKVRDEASSKANYDIDLDSGFDLDGRWVEIARNAEAGLGDVTKLQMRQLDLATKTKI 300
Q+VRD+AE +ANY+I+LDISGFD+DGRW+ETA+ NAEAGL DVI QMRQLD +TUK+

55 Sbjct: 241 QQVRDEASSKANYDIDLDSGFDLDGRWVEIARNAEAGLSDVITFKMRQLDFTKTKV 300

Query: 301 NGVVISNPPYGERLLDDKAVDILYNEMQQTFAPLKTNWKFILTSDEGFEKKYGSQADKKR 360
NGV+ISNPPYGERLLDDKAVDILYNEMQQTFAPLKTNWKFILTSDE FE KYG +ADKKR

60 Sbjct: 301 NGVVISNPPYGERLLDDKAVDILYNEMQQTFAPLKTNWKFILTSDELPACYGQADKKR 360

Query: 361 KLYNGTLLKVDLYQYGERVRRQV 383
KLYNGTLLKVDLYQ+YGERV+R +

65 Sbjct: 361 KLYNGTLLKVDLYQYGERVRRKL 383

SEQ ID 2966 (GBS255) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 43 (lane 7; MW 44kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 48 (lane 4; MW 69kDa).

-1071-

Sbjct: 229 ---DSDSNSSSSSSSSSD-SSDSDDSDSDSSSSSSSSSDS-SSSSSSSSSSSSSSSD 283

Query: 298 AVVDGKKSEKLEVDKGANFDSLSKTLNTGNASLDSLLHSIVSTRNQVKQSEBQASSNK 357
 + S+ + D+N S DS + + S DS S + N S+ S+
 Sbjct: 284 SSSSSSSSSSSSSSSSSN--SSSDSDSDSDSDSSSSSSSSSSSSSSSSSSSSSSSSSSS 341

Query: 358 VSDTQITQPNVINGQSSSSAATTNNQAAGTASGNLERNS 398
 + + + + + S+ S+ + N+ + + + + S
 Sbjct: 342 SSSSSSSSDSDSDSDSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 362

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2971> which encodes the amino acid sequence <SEQ ID 2972>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood =-14.70 Transmembrane 180 - 196 (175 - 202)

----- Final Results -----
 bacterial membrane --- Certainty=0.6880(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF15293 GB:AF202180 erythrocyte membrane-associated giant protein antigen 332 [Plasmodium falciparum]
 Identities = 41/173 (23%), Positives = 87/173 (49%), Gaps = 10/173 (5%)

Query: 1 VSEBSKEVEVKESQTLGLNAKSMITGEAVRKQSE---TKAGVTYKDSILDYIKQHR 56
 + E + V + K+ + GL+ + + +V+ +Q+E I + K+ S+ + +
 Sbjct: 78 IEEAEENVWIEKEVEEGLNREEVIDEEDSVSEQAEVEVINEEIKLKESSVDENVKVE 137

Query: 57 ---DEVSSQKFDKAYTELDASLNFIFKKQREALSKAGLVDPEVASAEQDSITLVEEV 113
 +EV+ + + LDN+ + + R+ + + +VD+ P S E E S + +EE+
 Sbjct: 138 LANEVNEETQSVARNNEEDKELNDVYVRETEGVTEVEVVDEVNSKVGQIEG-IIIEI 196

Query: 114 AEDLAPMETTAVVTGIPVEATVPVLDLDPSEVPIPEQMTEKPEKQDQLSED 166
 ED + G +E V + D SE + + E +T+E K+ + +ED
 Sbjct: 197 VEDGLTTDDLVOQQSGVIEVEVEVGS-D-SGIVREASITEEVEKKES-VTED 247

An alignment of the GAS and GBS proteins is shown below.

Identities = 234/506 (46%), Positives = 304/506 (59%), Gaps = 36/506 (7%)

Query: 1 MSDDKHPPFFEPKKETDGLFKDAKEMTVSEAVRKDSSEIKAGITEEDSILDKYIKQHR 60
 +SE+ K E KE+ L +AK MT+ EAVRK SEIKAG+T+DSILDKYIKQHR
 Sbjct: 1 VSEBSKE---VEVTKESQTLGLNAKSMITGEAVRKQSEIKAGVTYKDSILDYIKQHR 58

Query: 61 VASQKFETKSSDFANLDTASLDLPFIKKQRELGAMIARELSKLLDINSVQEQDTANAV 120
 V+SQKF+ K + LDASLD+FIKKQRE LG A + + + S EQ+
 Sbjct: 59 VSSQKFDK---YTELDASLNFIFKKQREALK---AGLVDDPEVASAEQDSITLVEE 112

Query: 121 SPKESSQEQNSVTVPPLNT-----EAPTEPEP---DSTIADSEYKSS 164
 + + + E VI +P T E + T EF D + + + + +
 Sbjct: 113 VAEDLAPMETTAVVTGIPVEATVPVLDLDPSEVPIPEQMTEKPEKQDQLSEDHHPAK 172

Query: 165 SKKRGIVGTLIALILLILVAITQYNYFQNNNSNTSQTATSSQSSSKATTSSSEDKKAS 224
 + G + L L+L + + +FG+N+P +S + S+ + + T S+ + + +
 Sbjct: 173 QNTKKGLIALILLILVAIVGNNHPLRQDSGKTTQTPASKQTKTSLQDTSAGKATFK 232

Query: 225 QNLDFNFKSYANFPVDDKKTLQNSSEFDKLSELEKVDALGTYGKVKVFKPSLRKQI 284
 F K Y F+ D K+LKNS F L +LE + AL+G+ YY K K K DSLK+ I
 Sbjct: 233 AAKAGFERLGTPTTNDKTLKNSGAFATLPDLAALKALEGSAAYDRAKAKVDSLRKAI 292

Query: 285 DAVKAVNKKFSPAVVDKKSEKLEKVDGANFDSLDKTLNTGNASLDSLLHSIVSTRN 344
 A+ AVN KF S VVDG+K EVK ANFD L S TL GNA+LD+L + + + GR

-1072-

Sbjct: 293 AAITAVNGKFPVSDVVVIGHEKVSA-EVKADANFDOLSSATLTIGNANLDAVLQASITESEQ 351

Query: 345 QVQCSSEEQASSNKVSDTQITEQPNVINGQSSSAATINNQAAGTAS--GNLERNRSVP 401

Q+ E A K ++ Q Q GQS+ S A + G S +L+R+ SRVP

Sbjct: 352 QLASKAEAA---KAANEQAV-QDQAAQGSTSVAFS---GYGLTSYDPAQLQRHLSRVP 403

Query: 402 YNNRAIADTGNPANI FNPQVLEKIVATSQARGYFSGNNYILPEVNI INKNGYNNMKLIG 461

YN IAD NP+W FNPQVLEKIVATSQARGY SGN YILEPVI INKNGYNNMK DG

Sbjct: 404 YNQDVIADRNANPNA FNPQVLEKIVATSQARGYISGNQYILEPVI INKNGYNNMKPDTG 463

Query: 462 TYLPSINAKITGYFVGNAPGRADSLDY 487

TYLPSIN KITGYFVGN G AD+LDY

Sbjct: 464 TYLPSINCKITGYFVGNKGYADALDY 489

SEQ ID 2970 (GBS351) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 2; MW 57kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 5; MW 82kDa).

GBS351-GST was purified as shown in Figure 216, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 972

A DNA sequence (GBSx1031) was identified in *S.agalactiae* <SEQ ID 2973> which encodes the amino acid sequence <SEQ ID 2974>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3169 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2975> which encodes the amino acid sequence <SEQ ID 2976>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3169 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 129/160 (80%), Positives = 149/160 (92%)

Query: 1 NTKRVVVEFELDHTIVKAPVRLISEEVGPVGDITINFDRLIQPNNAIDTAGLHTIE 60

NTKRV+VESFELDHTIVKAPVRLISEE GP GD ITNFD+RL+QPN+N+I+TAGLHTIE

Sbjct: 1 NTKREVESFELDHTIVKAPVRLISEEFGPKGDRITNFDVRLVQPNQNSIETAGLHTIE 60

Query: 61 HLLAKLIRQRINGLIDCSFPFGCRGTGPHIMWGKQDYTEIAKVIKSSLEAIGQVTWEDVP 120

HLLAKLIRQR+G+IDCSFPFGCRGTGPH+IMWGK +T+IAKVIKSSLE IA G+TWEDVP

Sbjct: 61 HLLAKLIRQRIDGMIDCSFPFGCRGTGPHIMWGKHSSTDIAKVIKSSLEIATGITWEDVP 120

Query: 121 GTTIESCGNYKDHSLSAQEWAKLILSQGISDNAPFRHIV 160

GTT+ESCGNYKDHSL +A+EWAKLIL+ QGISD+ P RH++

Sbjct: 121 GTTIESCGNYKDHSLPAKEWAKLILDQGISDPPSRHVI 160

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 973

A DNA sequence (GBSx1032) was identified in *S. agalactiae* <SEQ ID 2977> which encodes the amino acid sequence <SEQ ID 2978>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GF:AAF34762 GB:AF228345 unknown [Listeria monocytogenes]
Identities = 302/532 (56%), Positives = 400/532 (74%), Gaps = 14/532 (2%)

Query: 4 IILAVVICALIGLIIGYVAISMOMKSSKEAELTILNAEQDAVDLRGKAETAEHIRKAAE 63
I + ++ +L+ LI+G V S+ KSS E+ RG AE+ E +K AE
Sbjct: 3 IAITISSLFLFVLGLVVGSLIFKSS-----TEKLAARAGTAEILVVRDAKKEAE 52

Query: 64 RESKAHQKELLLEAKEEARKYREIEIEKFKSDRQELKQAEARLTFRASSLDKRDENLSNK 123
+KR LLEAKEE + R EIE E + R H ++ E RL R +LDKRD +LS +
Sbjct: 53 TT-----KKEALLLEKEEENHRLKTEIENELRGRRRTQKAENLLQREENLDRKDTLSKR 108

Query: 124 EKMLDSKQSLTDKSRHINERQEIATLETKKVEELSRIAEISQERAKDII LAOTEKDLA 183
E L+ KE+S++ + + I E+E ++A + + EL RI+ LS+EAQ IIL E++L
Sbjct: 109 EATLERKEESISGRQQOIEEKESKLAENIQABQTELERIGALSKEERAKSIIINQVEEILT 168

Query: 184 HDIATRIKRAERFVKDRSNKIADKDLAQAQCRILAGEYVTEQITITVHLPDDNMKGRIIGR 243
HD A +KE+E K+S+K AK++L+ A+QR A ++V E T++ V LP+D MKGRIIGR
Sbjct: 169 HDTAIVKESENRAKESBDDKKAIVLSLAIQRCADHVAETTVSVVLTENDMKGRIIGR 228

Query: 244 EGRNIRITLESITGIDVIIIDTPEVVVLSGFDPIRREIARMTLESITQDGRHIEAEELV 303
EGRNIRITL+L+TGID+IIDDTPE V+LSGFDPIRREIAR+ LE L+QDGRHIEAEELV
Sbjct: 229 EGRNIRITL+L+TGIDLIIDDTPEAVLSGFDPIRREIARLEKLVQDGRHIEAEELV 288

Query: 304 EKNRLEMDQRIREYGEAAAYEIGAPNLFHDLIKIMGRQFRTSYQGNVLRHSVEVGKLAG 363
+K R E+D+ IRE GE A +E+G ++HFDLIK+GRL++RTSYQGNVL HS+EV KLAG
Sbjct: 289 DKAREVDHIREYGEQATFEVGIHSHIPDLIKILGRVRYTSYQGNVLRHSVEVGKLAG 348

Query: 364 ILAELGENDLARRAGFLHDMGKAIDREVEGSHVEIGMEFARKYKEHPVMTVTIASHIG 423
ILAELGE+V LA+RAG LHD+GKALD+E+EGSHVEIG+E A KYKE+ ++N+IASHIG
Sbjct: 349 ILAELGENDVTLAKRAGLLHDIGKALDHEITGSHVEIGVELATKYKENDVVTINIASHIG 408

Query: 424 DVEPDVIAIVAAADALSSARPGARNESMENYIKRLRLLEEITANGFEGVQNAFALQAGR 483
D E SVIATV+VAADALS+ARPGAR+E++RNYI+RL LERL+ ++GV+ ++A+QAGR
Sbjct: 409 DTEATSVIAVLVAADALSARPGARSETLENYIRLEKLEISSESYDGVKSVIAQAGR 468

Query: 484 EIRIMVQPKGVSDDDQVIMSHKVEKIEQNLDPGNIKVTVIREMRVDFAK 535
E+RI+V+P + D ++ +R++IE+ LDYPG+IKTVIRE RAV++AK
Sbjct: 469 EVRIIEPDTIDDLSSYIRLARDIRKRISBELDYPGHKVTVIRETRAVEYAK 520
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2979> which encodes the amino acid sequence <SEQ ID 2980>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
```

-1074-

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has homology with the following sequences in the databases:

>GP:AAF34762 GB:AF228345 unknown [Listeria monocytogenes]
 Identities = 299/534 (55%), Positives = 408/534 (75%), Gaps = 14/534 (2%)

Query: 2 VNITLLIVSALIGLIGYALISIRLSAKRAAELTLINABQEAVIDIRGAQVADSHIRK 61
 + I + I+S+L+ LI+G + S+ KS+ B++ RG AB+ I +
 Sbjct: 1 MTAITITISSLLFLVGLVVGSLIPKES-----TEKKLAAAGTARL-----IVED 46

Query: 62 AKRESKANRKKLLLEAKERARKYREIEBOEFKSERQRLAQLETRLAERSITLORDENLS 121
 AK+E++ +KE LLEAKEE +R EIE E+ + R E++ E RL +R LORND +LS
 Sbjct: 47 AKKEASTTKGALLAKEENIRLATEIENELRGRTETQKARNLLQREENLORDNTLS 106

Query: 122 SKEKVLDSKEQSILTKSHIDERQLQVEKLSEKKAEHLKVAAMTARAREVILMETEN 181
 +E L+ KE+S++ + I+E++ ++ ++ ++ ELE++A++ EA+IL+ E +
 Sbjct: 107 KRATLERKEESISKRQQQITBEKSKLAEMTQAEYTELRISALSKEEASITLQVSEE 166

Query: 182 LTHEIATRIDAERDIKRTVKTAKDLLAQMQRLAGEYVTEQTITSVHLPDNNMGRII 241
 LTH+ A +++++E K+ + K AK++L+ A+QR A +IV E T++ V LP+D MGRRII
 Sbjct: 167 LTHDTAIMVKESENRAKESDCKARNILSLAIQRCAADHVAETTVSVTLFNDGNGRII 226

Query: 242 GRBGRNIRTLESITGIDVIIDDTPEVVILSGFDPPIRREIARMTLESILADGRHPIARIE 301
 GRBGRNIRTLE+LTDID+IIDDTPE VILSGFDPPIRREIAR+ LE L+ DGRHPIARIE
 Sbjct: 227 GRBGRNIRTLETILGIDLIIDDTPEAVILSGFDPPIRREIARILEKLVDGRHPIARIE 286

Query: 302 LVEKNRLEMONRIRETGAAAYEIGAPNHLFDLIKINGRLQRTSPQGNVLRHSVFGKL 361
 +V+K R E+D IRE GE A+E+G ++HFDLIKI+GRIL+RTS+QGNVL HS+EV KL
 Sbjct: 287 NVDKARKEVDHIREVGEQATFVSGIHSHPDLIKILGRILRYRTSYQGNVLRHSLEVSKL 346

Query: 362 AGILAGELGEVIALARRAGFLHDWKAIDREVEGSHVEIGMEFARKYKEHPVNTIASH 421
 AGILAGELGE+V IA+RAG LHD+GVAID E+RGSHVTEIG E A KYKE+ VV+N+IASH
 Sbjct: 347 AGILAGELGEDVTLAKRAGLLHDIGKAIDHEITGSHVEIGVELATKYKENDVIMTASH 406

Query: 422 HGDVSPDSVIAVLVAADALSARPGARMSMENYIKRLDLEIATSPDGVQNSPALQR 481
 HGD E SVIAVLVAADALSARPGAR+E+ENYI+RL LEEI+ S+DGV+ SA+QA
 Sbjct: 407 HGDVATSVIAVLVAADALSARPGARSETLENYIRRLKLEISESYDGVESYAIQA 466

Query: 482 GREIRIMVQPEKISDDQVILSHVREKIRNNLIYPGNIKVTVIREMRAVDYAK 535
 GRE+RI+V+P+ I D L+ +R++IE LDYGV+IKVTVIRE RAV+YAK
 Sbjct: 467 GREVRIIVEPDTIDOLSSYRLARDIRKKIEBELDYPGHIKVTVIRETRAVEYAK 520

45 An alignment of the GAS and GBS proteins is shown below.

Identities = 451/535 (84%), Positives = 503/535 (93%)

Query: 1 MFNIIAMVCAIGLIGYVAISMNMKSSKRAAELTLINABQEAVIDIRGAQVADSHIRK 60
 M NII L +V ALIGLI+GY IS+++KS+KEAAELTLINABQ+AVD+RGKAB++ASHI+K
 Sbjct: 1 MFNIIITLVSAIGLIGYALISIRLSAKRAAELTLINABQEAVIDIRGAQVADSHIRK 60

Query: 61 AARESKAHQKELLFAKEERARKYREIEBKESKDRQLQMQEARLITDRASSLORDENL 120
 A+RESKA++KELLLSAKERARKYREIE+EKES+RQLAQHE RL +R +LORDENL
 Sbjct: 61 TAKRESKANRKKLLFAKEERARKYREIEBOEFKSERQRLAQLETRLAERSITLORDENL 120

Query: 121 SNKEHMLDSKQSILTKSHIDERQLQVEKLSEKKAEHLKVAAMTARAREVILMETEN 180
 S+KEK+LDSKQSILTKS+HI+R++ ++ LE +K EL ++A++ ER++++IL +TE
 Sbjct: 121 SSKKKVLDSKQSILTKSHIDERQLQVEKLSEKKAEHLKVAAMTARAREVILMETEN 180

Query: 181 DLADHIAIRIKEASRVKDRSNKIADKLLAQMQRLAGEYVTEQTITVVHLPDNNMGRII 240
 L H+IATRI++AER++KDR+ K AKDLLAQMQRLAGEYVTEQTIT+VHLPDNNMGRII
 Sbjct: 181 KLTREIATRIDAERDIKRTVKTAKDLLAQMQRLAGEYVTEQTITVVHLPDNNMGRII 240

Query: 241 IGRBGRNIRTLESITGIDVIIDDTPEVVILSGFDPPIRREIARMTLESILADGRHPIARIE 300
 IGRBGRNIRTLESITGIDVIIDDTPEVV+LSGFDPPIRREIARMTLESILI DGRHPIARIE
 Sbjct: 241 IGRBGRNIRTLESITGIDVIIDDTPEVVILSGFDPPIRREIARMTLESILADGRHPIARIE 300

-1075-

Query: 301 ELVEKNRLMDQRIREYGEAAAYEIGAPNLHPDLIKIMGRILQPTTSYGQNVLRHSVEVGK 360
 ELVEKNRLMD RIREYGEAAAYEIGAPNLHPDLIKIMGRILQPTTSYGQNVLRHSVEVGK
 5 Subjct: 301 ELVEKNRLMDNRIRREYGEAAAYEIGAPNLHPDLIKIMGRILQPTTSYGQNVLRHSVEVGK 360

Query: 361 LAGILAGELGENVLDARRAGFLHDMKKAIDREVBGSHVEIGMEFARKYKEHPVVTIAS 420
 LAGILAGELGENV LDARRAGFLHDMKKAIDREVBGSHVEIGMEFARKYKEHPVVTIAS
 10 Subjct: 361 LAGILAGELGENVALARRAGFLHDMKKAIDREVBGSHVEIGMEFARKYKEHPVVTIAS 420

Query: 421 HHGDEVPCSVIAVVAADALSSARPGARNESMENYIKRLDLREIANGFEGVQNAFALQ 480
 HHGDEVPCSVIAV+VAADALSSARPGARNESMENYIKRLDLREIANGFEGVQNAFALQ
 20 Subjct: 421 HHGDEVPCSVIAVVAADALSSARPGARNESMENYIKRLDLREIANGFEGVQNAFALQ 480

Query: 481 AGREIRIMVQPGKVSDDQVIMSHKVRKTIQNLDYFGNIKVTVIREMRAVDFAK 535
 AGREIRIMVQPG KVSDDQVIMSHKVRKTIQNLDYFGNIKVTVIREMRAVDFAK
 15 Subjct: 481 AGREIRIMVQPEKISDDQVILSHKVRKTIQNLDYFGNIKVTVIREMRAVDYAK 535

SEQ ID 2978 (GBS86) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 6; MW 59kDa). It was also expressed in *E.coli* as a GST-fusion product.

20 SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 5; MW 84kDa).

GBS86-GST was purified as shown in Figure 192, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 974

25 A DNA sequence (GBSx1033) was identified in *S.agalactiae* <SEQ ID 2981> which encodes the amino acid sequence <SEQ ID 2982>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4984 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

35 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 975

40 A DNA sequence (GBSx1034) was identified in *S.agalactiae* <SEQ ID 2983> which encodes the amino acid sequence <SEQ ID 2984>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.87 Transmembrane 146 - 162 (146 - 162)

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.2147 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

50

-1076-

A related GBS nucleic acid sequence <SEQ ID 8697> which encodes amino acid sequence <SEQ ID 8698> was also identified. Analysis of this protein sequence reveals the following:

```

5      Lipop: Possible site: -1      Crend: 9
      MoG: Discrim Score: -10.72
      GvH: Signal Score (-7.5): -5.66
      Possible site: 29
      >>> Seems to have no N-terminal signal sequence
      ALOM program count: 1 value: -2.87 threshold: 0.0
10     INTEGRAL Likelihood = -2.87 Transmembrane 138 - 154 ( 138 - 154)
      PERIPHERAL Likelihood = 3.76 51
      modified ALOM score: 1.07

      *** Reasoning Step: 3

15     ----- Final Results -----
           bacterial membrane --- Certainty=0.2147(Affirmative) < succ>
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20     The protein has homology with the following sequences in the GENPEPT database.
      >GP:AG21390 GB:AF302051 ABC transporter ATP binding subunit
      [Bacillus licheniformis]
      Identities = 84/218 (38%), Positives = 138/218 (62%), Gaps = 1/218 (0%)

25     Query: 12 DIKVDHIFKSIQKTILEDISFSIASNQVALIGPWAGKTTLMSTLGLDISISSGSLT 71
      +++ + ++ K+ QKT ++ I FSI + VA++GPIGAGKTT +S +LG + ++G++T
      Sbjct: 3 NVVSLINVTIKTFRQKTAVDQIDFSIKKGSTVAILGPIGAGKTTITSMILGLLKPTAGNIT 62

      Query: 72 IFNLPAAHNRLKYKVAILPQR-NVLPSPKFTVRELIDPQRCLFPEVLPMSLILDYLNQSDT 130
      +F+ H R++ K+ + QE +V+P E+I+ R +P+ L + +D
30     Sbjct: 63 LFDSPHPEKRVREKIGITMLQEVSVMPGLRCRVEIILRSYYPKFLSFQKLRLTLTGLDK 122

      Query: 131 HLQQPTETLSGGQKRLLAFLVILVGVKQQLFLDEPTSGMDTSTRQRFWELIATLKKEGVT 190
      L+ E LSGGQKR L F L L G P+L+ DEPT GMD ++R RFW+ + +L ++G T
35     Sbjct: 123 DLKTQAEKLSGGQKRRLLGFALALAGDPELMIFDEPTVGMIDTSRNRFWQVQSLAEQGGK 182

      Query: 191 IVYSSHYIEVEHTADRILVLHKGKLLRDTTPELCHEAR 228
      I++S+HY++E + A RIL+ GK++ D TPL ++R
40     Sbjct: 183 IIFSTHYLQADDAQGRILLFKDGKIVADGTPLQIKSR 220

```

There is also homology to SEQ ID 686.

SEQ ID 8698 (GBS350) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 13; MW 28.9kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 4; MW 54kDa).

45 GBS350-GST was purified as shown in Figure 226, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 976

50 A DNA sequence (GBSx1035) was identified in *S. agalactiae* <SEQ ID 2985> which encodes the amino acid sequence <SEQ ID 2986>. Analysis of this protein sequence reveals the following:

```

      Possible site: 35
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
55     bacterial cytoplasm --- Certainty=0.2913(Affirmative) < succ>

```

-1077-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 977

- 10 A DNA sequence (GBSx1036) was identified in *S.agalactiae* <SEQ ID 2987> which encodes the amino acid sequence <SEQ ID 2988>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -10.51 Transmembrane 222 - 238 (214 - 241)
 15 INTEGRAL Likelihood = -6.90 Transmembrane 104 - 120 (101 - 125)
 INTEGRAL Likelihood = -5.84 Transmembrane 140 - 156 (138 - 159)
 INTEGRAL Likelihood = -5.20 Transmembrane 19 - 35 (18 - 41)
 INTEGRAL Likelihood = -1.28 Transmembrane 164 - 180 (164 - 180)
 ----- Final Results -----
 20 bacterial membrane --- Certainty=0.5203 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP: CAB69806 GB:A7243712 YVFS protein [Bacillus cereus]
 Identities = 73/239 (30%), Positives = 127/239 (52%), Gaps = 4/239 (1%)
 Query: 9 KMEFLITKRLANLINAIGMPVAFLLFFSGFMGEGLTKAEIYVRNMYITWAGFSSISF 68
 K+E L T R + ++ MPV F+ F+ + + +Y+I+NA FS +
 30 Sbjct: 4 KIELLRTFRNKLFIFFSLMPVMFYIIFTNVQ---VPQNGDANKAHLISNATFSIVGT 60
 Query: 69 AFTTFFPSMKDDQLSNRMQLLRHSFVPMWQYYLAKIRILFYCLAITVVFLLTHILRQV 128
 A F+P + ++ LL+ +P+P Y AKII +I V+P+ G++ +
 35 Sbjct: 61 ALPSFGVRLSQERQQWTHLLKITTLPPEGAYLTAKIQAQTVVNAFSLIVFIPTAGILINHV 120
 Query: 129 SMPIEQWQSFLLLLGGATCFIPFGLLVSYFKNTLSMSVANIYCYSMLAVLOGMNMPTM 188
 + I QN+ + L LL G T F+ G++ K + + +ANI MSLR++GG+NMPT +
 Sbjct: 121 ELTIQWIGAGLMLLLGVTFPLALGTVIGSIKKADAAAGLANILNMSLAIVOGLANMPIEV 180
 40 Query: 189 PFKWLQALSKLTPTTHTLVTLSPFANSFAGF-SLIILIGYGIIMLVIAVYLLSKRHSI 246
 PPK L+ + + TPTTH A G+ ++ +L GY +I +V++ + +++ ++
 Sbjct: 181 PPKILRTIGEWPTPTTFPGSGAWDIVAGKSIQWENTAVLOGYFLIFVVSIVYIRKQSAV 239

There is also homology to SEQ ID 682 and to SEQ ID 1628.

- 45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 978

- A DNA sequence (GBSx1037) was identified in *S.agalactiae* <SEQ ID 2989> which encodes the amino acid sequence <SEQ ID 2990>. This protein is predicted to be histidine kinase. Analysis of this protein
 50 sequence reveals the following:

Possible site: 49
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -7.43 Transmembrane 105 - 121 (102 - 124)

-1078-

INTEGRAL Likelihood = -6.95 Transmembrane 130 - 146 (129 - 149)

----- Final Results -----

bacterial membrane --- Certainty=0.3972 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9537> which encodes amino acid sequence <SEQ ID 9538> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAH54584 GB:AJ006400 histidine kinase [Streptococcus pneumoniae]
Identities = 138/350 (39%), Positives = 212/350 (60%), Gaps = 3/350 (0%)

Query: 11 MYFIPWFLYPIGGILYYHYPPWLFETLAFVGAYLYSVIRGSKYHMIANSTMLTY 70
M++I L+F+I+PI ++ W L + FV AY L V+ + + W MLTY+

15 Sbjct: 11 MFWISLIRMFIPILSVVTGWSAMHLLIDILFVAYL-GVLTTSKQLSLWLYGLMLTYV 69

Query: 71 FYMTIFNSGFNIWYIFLNNLLVYPRDK-LKSPRFISFACTLATVVF-LCFKASDPG 128
T F+ +IW+ +FLENLL Y P + LKS +F VV L F+ +

20 Sbjct: 70 VGNIAFVAVNYINWFFFLSNLSTHFEVRSLSLHVVTFLAQVLVVQQLIFQRIEVEP 129

Query: 129 RIMFLIVFICPGYMMIAIENRNSEQRKIAQNOYINILSAENERNIRGRDLHDSLGH 138
L++ F + + R E+ +E +QN IN+L AENR+RIG+DLHDSLGH

25 Sbjct: 130 LFYLLVILTFVDLMTFGELVRIVEDLKEAQVKQNAQINLLAENRSRIGQDLHDSLGH 189

Query: 189 TFAMTLKTRALKLEKRNVDKVCLESLNHIHQSGMSEVRQIVSNLKYRTVVEIDE 248
TFAM+++KT+IAL+L + Y +V+KEL E++ IS SM+EVR IV NLK RT+ E++

30 Sbjct: 190 TFAMLSVKTDLALQLQCMZAYPQVEKLEIKHQSISKDSHNEVRTIVENLKSRTVLSLET 249

Query: 249 LYRLQFLENIKLTVVKNLETSQSPVTOITIMILKELSNINIKHRAEADSVSLVRQGA 308
+ + + + I+ V N L+ S L+ +ST +MIL EL NI+KHA+ V L L R

30 Sbjct: 250 VKQMLEAGIEVQVENHLDKSLTQLESTASMLLELVNIIKHAKASVYLKLEKTEK 309

Query: 309 TINIMEIDNGGFINLDGDELHSIQERITVETLTILRSKPTHQVVL 358
+ + + D+GCGF ++ GDELH+++ R+ G ++++S+ PT +QV L

35 Sbjct: 310 ELILTVRDDGCGFASISGDELHVTNRVFFPSGEVSVISQKHPTVQVRL 359

There is also homology to SEQ ID 2992.

A related GBS gene <SEQ ID 8699> and protein <SEQ ID 8700> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8

McG: Discrim Score: 10.90

GVH: Signal Score (-7.5): -2.42

Possible site: 49

45 >>> Seems to have a cleavable N-term signal seq.

ALOM program count: 2 value: -7.43 threshold: 0.0

INTEGRAL Likelihood = -7.43 Transmembrane 105 - 121 (102 - 124)

INTEGRAL Likelihood = -6.95 Transmembrane 130 - 146 (129 - 149)

PERIPHERAL Likelihood = 0.16 61

50 modified ALOM score: 1.99

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.3972 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 979

A DNA sequence (GBSx1038) was identified in *S.agalactiae* <SEQ ID 2993> which encodes the amino acid sequence <SEQ ID 2994>. This protein is predicted to be response regulator. Analysis of this protein sequence reveals the following:

```

5  Possible site: 28
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.16    Transmembrane    49 - 65 ( 49 - 65)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:CAB54585 GB:AJ006400 response regulator [Streptococcus pneumoniae]
      Identities = 95/153 (62%), Positives = 125/153 (81%), Gaps = 3/153 (1%)

Query: 1  MKLLVAEDQSMRLDAMCOLLLMESVSTTDQAGNGGEAIAILSNKAIDVAILDVENPILS 60
      ME+LVAEDQSMRLDAMCOLL+++ V ++ QA NG EAI +L +++D+AILDVEMP+ +
20  Sbjct: 1  MKVLVAEDQSMRLDAMCOLLLMQPDVESVQAQNGGEAIAQLLEKESVDIAILDVENPVKI 60

Query: 61  GLDVLEWVRKYQ-MVKVIIITTFKRSQYFQRAIRSNVDAYVLKDRSVADLWATIKVLGSG 119
      GL+VLEW+R + KV++VITTFKR GYF+RA+++ VDAYVLE+R++ADLM+T+ VL G
25  Sbjct: 61  GLEVLWIRAEKLETKVTVVVTTFKRPQYFERAVKGVDAIVLKERNIADLMQTILHVLGSG 120

Query: 120  GKEYSPELMENV--SNPLSSQETIKILSLIAQG 150
      KEYSPELME V+ NFL+EQEI +L IAQG
      Sbjct: 121  RKEYSPELMENVMMHFNPLTQELIADVLEGLIAQG 153

```

30 There is also homology to SEQ ID 2996.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 980

A DNA sequence (GBSx1039) was identified in *S.agalactiae* <SEQ ID 2997> which encodes the amino acid sequence <SEQ ID 2998>. Analysis of this protein sequence reveals the following:

```

35  Possible site: 34
   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -6.69    Transmembrane    156 - 174 ( 145 - 184)
      INTEGRAL    Likelihood = -4.94    Transmembrane    11 - 27 ( 8 - 31)
40  INTEGRAL    Likelihood = -3.93    Transmembrane    74 - 90 ( 73 - 92)
      INTEGRAL    Likelihood = -2.39    Transmembrane    103 - 119 ( 102 - 119)
      INTEGRAL    Likelihood = -1.86    Transmembrane    42 - 58 ( 38 - 59)

45  ----- Final Results -----
      bacterial membrane --- Certainty=0.3675(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

50  >GP:AB88565 GB:AB000909 unknown [Methanothermobacter
      thermoautotrophicus]
      Identities = 46/183 (25%), Positives = 81/183 (44%), Gaps = 11/183 (6%)

Query: 5  KERFDYLSDAIATAMTILVLEI-----KTPATMGDIGDFTRNIGLFVSVVVVFW 57
      K+R + L DAI AIAMTILVL I PA I ++ + +SP+++ FW
55  Sbjct: 6  KRLGLSLVDAIFAAMTILVLCIDVPTGTMSEVPAMDAYIMGLASDLYSYCLSPILLGVFW 65

```

-1080-

Query: 58 YERAQNSLDAQKTNDIEIALDIEHLGICLIPLFTKPMISPEHNINFAVMAYGLLTLVGL 117
 + + +K + I ++I+ + + L+P TK ++ + + L L +GL
 Sbjct: 66 WVMHMFEEKLEKVDGTGFWINIVLWVVLVFPSTKLGTNGVGLVTNLFHLMGTGL 125

Query: 118 TSDIIRIRLASDYDLVITPSELKERVKMTTFAIRSVVVRFTIILAYLFEVGFIFAYLV 177
 + I L+ I ++K ++ + +L PE AY V
 Sbjct: 126 LLSMSWITQTRNGIMDIGRNEYRLILKKNLMLPLAI----LALITPTIAPYBSSTAYAV 181

Query: 178 IPL 180
 + L
 Sbjct: 182 LIL 184

No corresponding DNA sequence was identified in *S.pyogenes*.

- 15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 981

- A DNA sequence (GBSx1040) was identified in *S.agalactiae* <SEQ ID 2999> which encodes the amino acid sequence <SEQ ID 3000>. This protein is predicted to be guanylate kinase (gmk). Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

25 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:CB13441 GB:Z99112 similar to guanylate kinase [Bacillus subtilis]
 Identities = 121/202 (59%), Positives = 155/202 (75%)

Query: 1 MSEGGLIVSPGSGVGKGTVRQEIFSTPDHKPFDYSVMITRPFQGEVVDGVYFFKTR 60
 M EGGGLIV SGPSGVGKGTVRQ IFS D KF+YS+S+TTR R GEV+GVDYFF+TR+
 Sbjct: 41 MKERGGLIVLSPGSGVGKGTVRQAIFSQEDTKPEYSIVTTRSPRSGEVMGVYFFKTRD 100

Query: 61 EPEALIKRQGMLEVAEYVGNYYGTPPLSYVNETLDKIDVFLTEVQCALQVSKVVDGVF 120
 EPE +I + ++LE+AEYVGNYYGTP+ YV +TL G DVFLTEVQCALQV++ P+G+G
 Sbjct: 101 EFEQMIADNKLLEVAEYVGNYYGTPVDYVEQLDQKDVFLTEVQCALQVRNAPFGLF 160

40 Query: 121 IFLTPDLEERLVRGRTDSEVLAQRTERAKREIALMREYDVAVVNDQVSLAAERVK 180
 IFL PP L EL+ R+V RGT++ +I R++ AK EI +M YDY V ND V A +++K
 Sbjct: 161 IFLAPFSLSELKNRIVIRGTETDALIENRMKAARIEEMQDAYDYVVRNDVETACKIRK 220

45 Query: 181 RVIAEHYRVDVRVIGRYNIMVK 202
 ++ AEH + +RV RY M++
 Sbjct: 221 AIVLAHLKREKRVAPRYKML 242

- A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3001> which encodes the amino acid sequence <SEQ ID 3002>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

55 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-1081-

The protein has homology with the following sequences in the databases:

>GP:CAB13441 GB:Z99112 similar to guanylate kinase [Bacillus subtilis]
Identities = 123/203 (60%), Positives = 157/203 (76%)

5 Query: 1 MSERGLLIVSPGSGVGKGTVRQEIFSTPDHKKFEYSVSMITRQPRGEVDGVDYFFRTRE 60
N ERGLLIV SPGSGVGKGTVRQ IPS D KPEYS+S+TTR R GEV+GVDYFF+TR+
Sbjct: 41 MSERGLLIVSPGSGVGKGTVRQAIIPSQEDTKFEYSISVTTSPRBEVNGVDYFFKTRD 100

10 Query: 61 EFEELIKIQGMLEYAERYVGNYYGTFPLVYVNETLDKGIDVFLIEVQCALQVSKVPDGVF 120
EER++I ++LE+AEYVGNYYGTF+ YV +TL G DVFLIEVQCALQV++ P+G+P
Sbjct: 101 EFEEQMIADNKILEAERYVGNYYGTFVDYVETLQDGRDVFLEIEVQCALQVRNAPFEGLP 160

15 Query: 121 VFLTPDLELEDRLVGRGTDSPVIAQRIERAKEEIALMREYDVAVVNDQVSLAERVK 180
+FL PP L EL++R+V RGT++ +I R++ AK EI +M YDY V ND V A ++K
Sbjct: 161 IFLAPPSLSELNRIIVRGSTETDALIENRMKAKEAETEMDAYTVVNDVETACDIK 220

20 Query: 181 RIETEHRFRVERVIGRYDKMIKI 203
I+ EH + ERV RY KM++
Sbjct: 221 AIVLAHLKRERVAFRYKIMLEV 243

An alignment of the GAS and GBS proteins is shown below.

Identities = 186/204 (91%), Positives = 197/204 (96%)

25 Query: 1 MSERGLLIVSPGSGVGKGTVRQEIFSTPDHKKFEYSVSMITRQPRGEVDGVDYFFRTRE 60
MSERGLLIVSPGSGVGKGTVRQEIFSTPDHKE+YVSMITRQPRGEVDGVDYFFRTRE
Sbjct: 1 MSERGLLIVSPGSGVGKGTVRQEIFSTPDHKKFEYSVSMITRQPRGEVDGVDYFFRTRE 60

30 Query: 61 EFEELIKIQGMLEYAERYVGNYYGTFPLVYVNETLDKGIDVFLIEVQCALQVSKVPDGVF 120
EFE LIK QGMLEYAERYVGNYYGTFPL+YVNETLDKGIDVFLIEVQCALQVSKVPDGVF
Sbjct: 61 EFEELIKIQGMLEYAERYVGNYYGTFPLVYVNETLDKGIDVFLIEVQCALQVSKVPDGVF 120

35 Query: 121 IFLTPDLELEDRLVGRGTDSPVIAQRIERAKEEIALMREYDVAVVNDQVSLAERVK 180
+FLTPDLE+EL+RLVGRGTDSPVIAQRIERAKEEIALMREYDVAVVND+V+LAERVK
Sbjct: 121 VFLTPDLELEDRLVGRGTDSPVIAQRIERAKEEIALMREYDVAVVNDQVSLAERVK 180

40 Query: 181 RVIEAEHYKVDRIYGRYDMVKET 204
R+IE EH+RV+RVIGRY M+K T
Sbjct: 181 RIETEHRFRVERVIGRYDKMIKI 204

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 982

A DNA sequence (GBSx1041) was identified in *S. agalactiae* <SEQ ID 3003> which encodes the amino acid sequence <SEQ ID 3004>. Analysis of this protein sequence reveals the following:

Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.1763 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3005> which encodes the amino acid sequence <SEQ ID 3006>. Analysis of this protein sequence reveals the following:

Possible site: 35
>>> Seems to have no N-terminal signal sequence

-1082-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1551(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 95/105 (90%), Positives = 100/105 (94%), Gaps = 1/105 (0%)

Query: 1 MMLKPSIDTLLDKVFSKYSVLVILQAKRAHLEAGEKATQDFKSVKSTLRALEHIESQNVV 60
 MMLKPSIDTLLDKVFSKYSVLVILQAKRAHLEAG TQ+FKSVKSTL+ALEHIESQNVV
 Sbjct: 1 MMLKPSIDTLLDKVFSKYSVLVILQAKRAHLEAGATPQRFKSVKSTLQALEHIESQNVV 60

Query: 61 IHFDPSAKRASVRARIFAERLAKKEERKIKEQIAKEK-EDGEKI 104
 IHFDPSAKR +VRA+IRARLAKKEERKIKEQIAKEK E+GEKI
 Sbjct: 61 IHFDPSAKRBAVRAKIFAERLAKKEERKIKEQIAKEKEEBGEKI 105

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 983

A DNA sequence (GBSx1043) was identified in *S.agalactiae* <SEQ ID 3007> which encodes the amino acid sequence <SEQ ID 3008>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3413(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13444 GB:Z99112 primosomal replication factor Y (primosomal protein N') [Bacillus subtilis]
 Identities = 377/807 (46%), Positives = 529/807 (64%), Gaps = 21/807 (2%)

Query: 6 AQTIVDIPLMQTKDEPPSYALPKDLEDLVQGVGVVIFPGRGNRLQGVFVGRDDDELET 65
 A+VIVD+ D+PF Y IP L+ + + + G+RV VFFG R+QGVF ++ +L
 Sbjct: 4 AETIVDVSTKNTDRPFDYKIPDLHKGMIKTGKGVIVFFGP- -RKIQGVFVAVKASDLSG 61

Query: 66 KDIAEV---LDPEFVLNQEQDLADQMRHTVFSYKISILKMSPLSLNYSVKLL---A 119
 K + EV LD PVL +E + L+ + S+KI+ L++MLP+ L ++Y+K L
 Sbjct: 62 KSVKEVEDLDDLTVPVLTEEMILSSWISDKTLSPKTLQALQMLPAALAKAYEKELKIAHG 121

Query: 120 TDYLPSEDRDLFGHKTRIVPSSLSQDAKKA-GRLLQGFPIEVQYLAOKKTKTKETIY 178
 D P +R LP +++S + + K R +QNG I+V Y K K + +
 Sbjct: 122 ADLPFQVQR--LFSETKTLNYSIDPDHETLKLQRHVQSGIDVTVKVAQKTNQGVH 179

Query: 179 KINRTLEKQS----IARAKKRIELKEFLLENPQGRITALN---KQSPSSVVNFVRE 230
 + N + E ++ ++ +A K+ + FL+ P+ + +A SS + +
 Sbjct: 180 QANASKEELAKQAGLSRQNAQQAIIHLFLISEPBGVKIPAEELCKKTITVSATITKTLIQ 239

Query: 231 BGIEVIEKEASRSDNYPKGILKTDFLDLNGQAKVVKVVDIQGKEQNKPFLLGIGTGS 290
 +G+++ +E R K KT+ L L EQ + + + + +++K FLL G+TGS
 Sbjct: 240 KGLLRSYEEVYRDYQDMFKKTEPLFLPDBQRAAFPEIKETLDSDEHVFLLHGVIGS 299

Query: 291 GKTEVYLIIIDNVILKLGKTAIVLPEISLITPQMTNRFISRPQKQVAIMHSGLSSEKEDR 350
 GKTE+YL I+ VL GK AIVLPEISLITPQMRNRFQGVQVA+MHSGLS GKE+DE
 Sbjct: 300 GKTEIYLGKIEKVLAKGKTAIVLPEISLITPQMVNRFPKGRFGSQVAVMHSGLSSTGKEDR 359

Query: 351 WRKIKSQAKVVGARSATFAPFENLIGATIIIDHEHSTYQBSNFYHARDVALLPASYV 410
 WRKI + ++VVGARSATFAP EN+G IIDHEHSTYQBS FRYHA++VA+ RAR++
 Sbjct: 360 WRKIRKEVRLVVGARSATFAPFENLIGATIIIDHEHSTYQBSFRYHAKVAIKRAH 419

-1083-

- Query: 411 KAVLLMGSA TPSSIESRARSRDVYKFLKLRHNPARKARIPOVEIIDFRNFIGQQEVSNT 470
 +++GSATP+ES ARA + VY+ L LKIR N + +P+V ++D R + S F+
 Sbjct: 420 SCPVVLGSATPFLSEYARAQKGVYELLSLKRVNHRV-MPEVSLVDMREELNGNRSF 478
- Query: 471 SYLLDKIRDRLDKKQVVLMMNRGYSFIMCRDCGSVDQCPNCDISLITLMAKMTKNC 530
 L++K+ + + K EQ VL LN+RGYSS+MCRDCGYV QCP+CDIS+T H + + CH
 Sbjct: 479 VETMEKLEETIAKGEQAVLFLKRGYSSFVMCRDCGSVVRQCPHCDISMTYRYQALKCH 538
- Query: 531 YCGFEKPIPRTPCNCNKSISYVGTGTQKAYEELLKVPIDAKILRMDVDTTQKGGHRSI 590
 YCG R+P+P TCP C S+ I ++GTGTQ+ EEL KV+P A+++RMDVDTT +XG HR+
 Sbjct: 539 YCGHEEPVPHTCPECASSHIRFFGTGTQVRVEELKVLSEARVIRMDVDTTSRKGAKHL 598
- Query: 591 LKRFGNHEADILLGTQMLAKGLDFPNVTLAVGLNADTSLNLPDFRESSERTFOLLTQVAGR 650
 L PG +ADILLGTQMLAKGLDFPNVTLAVGL+ADT+L++PDFRS+R+TFOLLTQV+GR
 Sbjct: 599 LSAFGBGKADILLGTQMLAKGLDFPNVTLAVGLSADTTIHLPDFRSAREKTQVQVGR 658
- Query: 651 AGRAEKKEGEVVIQTYNPMHYAIQLAQKQDFEAFYQYMNIRROLGYPPYFTVGLTSHK 710
 AGR EK G V+IQTY P+HY+IQL + D+E FYQ+EM RR+ YPFY+ +T+SH+
 Sbjct: 659 AGRHEKPGHVIQITPSSHYSIQLTKTHDYETFYQHEMAHREQSYPPYTYIALVTRVSH 718
- Query: 711 DREWLIRKSYEVLGLLQKQGSQKVLKSGPTPKPIARTHNLVHYQILIKYRFEDNLEVLN 770
 + ++ ++ LK K+LGP+ PIAR + Y YQ +IKY+ E L +L
 Sbjct: 719 EVAKAAVTAEKIAHFLKANGADTKILGPSASPIARIKDRYQCVIKYQSTQLSALK 778
- Query: 771 RLID-MTQDKENRDLRLAIDHFQNM 796
 ++L+ ++ E + + ++ID P MM
 Sbjct: 779 KILSHYKRELEQHVMSIDMNPYMM 805
- 30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3009> which encodes the amino acid sequence <SEQ ID 3010>. Analysis of this protein sequence reveals the following:
- Possible site: 32
 >>> Seems to have no N-terminal signal sequence
- 35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1396 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
- 40 An alignment of the GAS and GBS proteins is shown below.
- Identities = 556/793 (70%), Positives = 659/793 (82%), Gaps = 1/793 (0%)
- Query: 4 KLAQVIVDIPLAQTDKPPSYALPKDLEDLVQGVGVHVPFGRGNLLQGFVVGPRDDEL 63
 K+A VIVDIPLAQTDKPPSY IPK+L LVQ+G RVEVPFG+GNRLQGF++GF +D
 Sbjct: 12 KVAHVIVDIPLAQTDKPPSYGIPKRLVSLVQLGRVHVPGKGNRLQGFIIIFGQGEDS 71
- Query: 64 ETKDIAEVLDFEVLNQBQLDADQMRHTVFSYKISILEKMLPSLLNSQYDKLLLATDTL 123
 K I VLD EPLVNQBQL IADQ+R TVFSYKI++LK++P+ILNS YDK+L L
 Sbjct: 72 SLKLIQTVLDFEVLNQBQLTADQLRKTVFSYKITLLKMLPNLNSQYDKLLLATDTL 131
- Query: 124 PSKDRDLFGHKTRIVFSSLSQAKKAGRLIQKQFIEVQYLARDKKTIKTEKIYKINRT 183
 DR+ LF K +++S+L + K A + IQ G I V YLARDKK +KTEK Y ++
 Sbjct: 132 KKSDRDLFGHKTRIVFSSLSQAKKAGRLIQKQFIEVQYLARDKKTIKTEKIYKINRT 191
- Query: 184 LLEKSQAARAKKRILEKSEFLNPPQGRITALNKPQSSPVNFFREGLIEVIEKEASR 243
 L I++RAKKR LK++L + + +L L + FS VV +F +I +B+ R
 Sbjct: 192 ALAVHPISSRAKKRQLKLYLLTHTKAKLATLXQAFSDVVAIFYVINLIRIDRPSIDR 251
- Query: 244 SDNYFKGLKTDPLDINQQAQVVKIVVDQIGKEQNKPFLEIGITGSGKTEVYLHIDNV 303
 S++VF I + FL LN++QA V +V+QIGK +KPFLEIGITGSGKTEVYLHIDNV
 Sbjct: 252 SRSYFDQIKPSSFLTNQQAQVATRIEVBQIGK-SKPFLEIGITGSGKTEVYLHIDNV 310
- Query: 304 LKLGKTAIVLVPRISLTQPMNRFISRFQKQVAIMHSGLSGKDFBMRKIK+GQAKVVV 363
 LK KTAIVLVPRISLTQPMNRFISRFQKQVAIMHSGLSGKDFBMRKIK+GQAKVVV
 Sbjct: 311 LKQDKTAIVLVPRISLTQPMNRFISRFQKQVAIMHSGLSGKDFBMRKIK+GQAKVVV 370

Query: 364 GARSALFAPLENIGAI IIDBESHSTYKQESNPYHARDVALLRAEYYKAVLNGSATPSI 423
 GARSALF+PLE IGAI IIDBESHSTYKQESNPYHARDVALLRAE+++AV+NGSATPSI
 Sbjct: 371 GARSALFSPLERIGAI IIDBESHSTYKQESNPYHAREVALLRAKHQAVVNGSATPSI 430

Query: 424 ESRARASRDVYKPLEKIRANPKARIPQVEIIDFRNFIGQQEVSNPTSYLLDKIRDLK 483
 ESRARAS+ VY F++L RANP A+IP+V I+DFR++IGQQ VSNPT YL+DKI++RL K
 Sbjct: 431 ESRARASGVYHFILQLTQRANPLAKIPEVTIVDFRDYIGQAVSNPTPYLIDKIKKRIK 490

Query: 484 KBQVVLMLNRRGYSSFMCRDQGVVDQCPNCDISLTHMATKTMNCHYCGFKEPIPTCP 543
 KBQVVLMLNRRGYSSFMCRDQGVVD+CPNCDISLTHM TKTMNCHYCGF+KEPI TCT
 Sbjct: 491 KBQVVLMLNRRGYSSFMCRDQGVVDKCPNCDISLTHMDTKTMNCHYCGFKEPIPTCP 550

Query: 544 NCNKSISYVYGIGTQKAYRELLKVIPDAKILRMDVDTTRQGGHESILKRPNGHREADILL 603
 C+S SI YVYGIGTQKA++EL VIP+AKILRMDVDTTR+K H++IL FG EADILL
 Sbjct: 551 ECHSNSIRYVYGIGTQKADDELQGVIPRAKILRMDVDTTRKKRSHTILDSFGREQADILL 610

Query: 604 GTQMIAGLDFPNVTLVGLVNADTSLNLPDPFRSSERTFOLLTVQVAGRAGRAKKGESVVIQ 663
 GTQMIAGLDFPNVTLVGLVNADTSLNLPDPFR+SE+TFOLLTVQVAGRAGRA K GSV+IQ
 Sbjct: 611 GTQMIAGLDFPNVTLVGLVNADTSLNLPDPFRASEKTFOLLTVQVAGRAGRAKKGESVVIQ 670

Query: 664 TYNPNHYAIQLAKQDQFAFYQYENIRRLQGYPPYPTVGLTSLHKDEEWLIRKSYEVL 723
 TYNP+HYAIQLA+QDQFAFY+YEN+IR Q+ YPPYPTVG+TSLH+ E +++K+Y+V
 Sbjct: 671 TYNPDHYAIQLAKQDQFAFYKYENIRRLQGYPPYPTVGLTSLHLEASVYKGAQVTC 730

Query: 724 SLLKQGSFSDKVLGGPTPKFIARTNLYHYQIIKRYFEDNLEVLNRLDMTDQENRND 783
 LLK SD +K+LGPTPKFIARTNLYHYQI++KYRFEDNLE LNR+LD +Q+ +NR
 Sbjct: 731 ELLKSHLSDNLIKLGPTPKFIARTNLYHYQIIKRYFEDNLEETINRLDMTQENRND 790

Query: 784 LRLAIDHEPCNM 796
 L+L ID EPQ +
 Sbjct: 791 LKLIIDCEPQQL 803

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 984

A DNA sequence (GBSx1044) was identified in *Sagalactiae* <SEQ ID 3011> which encodes the amino acid sequence <SEQ ID 3012>. This protein is predicted to be methionyl-tRNA formyltransferase (fmt). Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1329 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAE13446 GB:Z99112 methionyl-tRNA formyltransferase [Bacillus subtilis]
 Identities = 155/314 (49%), Positives = 221/314 (70%), Gaps = 7/314 (2%)

Query: 1 NTKLLFMGTDPDSATVLKGLIADQKYDVAVVTQPDRAVGRKKEIKMTPVKEVALENNIP 60
 NT+++FMGTDPDS VI++ DO Y+V+ VVTQPD R GRKK + PVKE AL + IP
 Sbjct: 1 NTRIVFMGTDPDSVPVRLTIEDG+YSVGVVTQPD RPKGRKVKVLTTPPVKSEALRHGIP 59

Query: 61 VYQPKLGSPELQMLTIGADGIVTAAQQLPTKLLBSVGFA-INVHASLLPKYRGA 119
 V QPK+ + E+E+++ L D IVTAAQQL LP +LL+S + INVHASLLP+ RGA
 Sbjct: 60 VLQPKVRLTEIEKVLAKPDLIVTAAQQLPKELLDSPKSGCINVASLLPELRGA 119

Query: 120 PIHYAINGEKEAGVTIMEMAKMDGDMVSKASVEITDENVGTIMFDRLAVVGRDLDD 179
 PIHY+I+ G+K+ G+TIM MV K+DAGUM+SK V+I + DNVGT+ D+L+V G LL +

-1085-

Sbjct: 120 PIHYSILQKKKIGITIMYVVKLDAGDMISKVEVDIEETDNVGLHDKLSVAGAKLLSE 179
 Query: 180 TLPGHLSGDIPQNEEEVSPSPNISDPDEERIDWNKSSRDIPNHVGMYPNPVVAHTLLE 239
 T+P ++G I P Q+EE+ ++PNI ++E +DN++ ++N +RG+ PNPVAT L
 5 Sbjct: 180 TVFNVIAGSISPEKQDEKATYAPNTRKREJELLONSRTGEELYNQIRGLNPNPVAITTLN 239
 Query: 240 GHRFKLY--EVTMSBGKSGSPQVIARTKNSLIVATG-DGAIELKSVQPAGEPRMDLKDPL 296
 G K++ + + PG V+A K + VATG + A+ L +CPAGK RM +DF+
 10 Sbjct: 240 GQNLKINWASKKIAAPTTPAFGIVVAVEKSGIIVATGNKTAALLTELQCPAGKRMKGEDFV 299
 Query: 297 NGVGRNLEIGDKFG 310
 G ++E GD G
 Sbjct: 300 RGA--HVENGDVLG 311

- 15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3013> which encodes the amino acid sequence <SEQ ID 3014>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0730 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 25 An alignment of the GAS and GBS proteins is shown below.

Identities = 217/310 (70%), Positives = 266/310 (85%)

Query: 1 MTKLLFMGTDFSFATVLKGLADGKIDVLAVVTPDRAVGRKKGEIKMTVPVKEVALENNIP 60
 M KLLFMGTDFSFATVLKGL+L + Y+LL VVTQDRAVGRKK+IK+TPVK++ALE+ I
 30 Sbjct: 1 MTKLLFMGTDFSFATVLKGLLDNFAEYILGVVTQDRAVGRKKDIKVTTPVKQALEHGIS 60
 Query: 61 YIQPEKLSSSPLEQLWTLGADGIVTARQQFLPTKYLLESVGFAINVHASLLPKYRGAP 120
 +YQPEKLSSS EL ++M LGADGI+TAAPGQFLPT LL+SV FAINVHASLLPKYRGAP
 35 Sbjct: 61 YIQPEKLSSQELIEIMELGADGIIITAPGQFLPTILLDSVSFAINVHASLLPKYRGAP 120
 Query: 121 IHVAIINKEKAGVTIMEMVAKMDAGDWKSKASVETIDEDNVGTMPDRLAVVGRDLLDIT 180
 IHVAI+NG+KEAGVTIMEM+ +MDAGDW+KAS I + DNVGT+P++LA++GRDLLD+
 40 Sbjct: 121 IHVAIMKGDKEAGVTIMEMIKEMDAGDWAKASTPILETQNVGTFLFKLAIIGRDLLDS 180
 Query: 181 LPGLYSQDIKPIQNEEEVSPSPNISDPDEERIDWNKSSRDIPNHVGMYPNPVVAHTLLE 240
 LP YLSQ+KPIQ+ + +FSFNISS+ E++DW E++++FNI+RGM PNPVAT LEG
 45 Sbjct: 181 LPAYLSGLKLP+PQDHSQNTFSFNISSPEHEKLDWMSQEVFNIHIRGMNPVVAHTFLEG 240
 Query: 241 NRKFLYEVMTSBEKSGSPQVIARTKNSLIVATGDGAIELKSVQPAGEPRMDIKDFLNKVG 300
 R K+YE ++EG+G PGQV+ KTK SL +ATG GA+ L VQPAGEP+M I DFLNG+G
 45 Sbjct: 241 QRLKIYEAQLABEGLPQGVVVTKKSLVIATQQAGLSLIVQPAGEPKMSIDIFLNGIG 300
 Query: 301 RNLEIGDKFG 310
 R LE+GD G
 50 Sbjct: 301 RKLKLVGDIIG 310

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 985

- 55 A DNA sequence (GBSx1045) was identified in *S.galactiae* <SEQ ID 3015> which encodes the amino acid sequence <SEQ ID 3016>. This protein is predicted to be sunL protein (sun). Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

-1086-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1677 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA10711 GB:A1132604 sunL protein [Lactococcus lactis]
 Identities = 222/434 (51%), Positives = 305/434 (70%), Gaps = 15/434 (3%)

Query: 7 KSARGALMTLEKVPKQYNSIALNKLKSRSLSDKRALVTEIVYGVTKITLWYL 66
 K+AR AL L ++F AY+NI+L++L+ S LS D+ VT +VYG V++K LEWY+
 Sbjct: 3 KNARQTALDVLNDIFGNDAVANTISLDRNLRSDELSTVDKGFVTALVYGVVSKALLEWYL 62

Query: 67 SHFIVDRDKLEINWYHLLLSLXQLLYLONIPDHAIVNDVATIAKRNKNGKGAELINAV 126
 + + K W LLLL++YQ+L++D +P A V++AV IAK R + + INAV
 Sbjct: 63 TPLKKKEPK--WAWMLLLLTYYQLFMDKVPISAAVDEAVKIAK-RHDGQATANFINAV 119

Query: 127 LRR-VSSETPLIASIKRQKRYSVAYSMPVWLVLKLLIDQYGTTRALAIMSLFERNKAS 185
 LR + SE E + K + YSMP L+ K++ Q+G R I+ESL + + S
 Sbjct: 120 LRNFMRSSEHRNE-----EPDQWETKYSMPKLLLDMDVVRQFGGKRTGEILESLEKSHVS 173

Query: 186 LRVTLSQKQKTIKETLVNDRSDIAETALVADSGNFASTSFQGLITIQDESSLQVAPL 245
 LR D + E R S + ETAL+ADSGNF+ T FQ G ITTIQE+SQLVAP L
 Sbjct: 274 LRKIDPTV-----SIAGTRPSLLTETALADSGNFSITTEPQIGRTIT:QDTSQLVAPQL 228

Query: 246 KVSNDQVILRCSAPGKTKSHIASYLTGAVTALDLYDHKLEINWMAKRIGLSDKIKTK 305
 ++ G ++VLDAC+APGKG++HA+ YLTGG +TALDLY+HKL+L+ +NA+R ++DKI T+
 Sbjct: 229 ELEGFEEVLDACAPGKGKTHMAQLTITGHTTALDLYEHKLLDINQNAQGHVADKTIQTQ 288

Query: 306 KLDASKAHEVFLKDTFKILVNDAPCSGIGLIRKPKDPIKYNANQDFALQRTQLSLGAV 365
 K DA+ +E F + FD+ILVNDAPCSGIGLIRKPKDPIKRYKSSDFIDLKTIQELINSA 348
 Sbjct: 289 KADATWYIENFGPEKFDRIILVNDAPCSGIGLIRKPKDPIKRYKSSDFIDLKTIQELINSA 348

Query: 366 QOTLRKGGIITYSTCTTIFEEHNPQVIEKFLNHNHNPQVLSHSDRIVKRGCTISLSPQ 425
 ++L+K GI+ YSTCTTIF+EEHNP V+ +FLENHNHNPQVIE+S+ + + + GC+ I+PE
 Sbjct: 349 SKELKSGIMVYSTCTTIFDEHNPVVEHFLNHNHNPQVIE+SNKPEVIKSCLETFPEN 408

Query: 426 YHTDGFPIQVKKRI 439
 YHTDGFPI + K+I
 Sbjct: 409 YHTDGFPIAKFKKI 422

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3017> which encodes the amino acid sequence <SEQ ID 3018>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA10711 GB:A1132604 sunL protein [Lactococcus lactis]
 Identities = 208/433 (48%), Positives = 287/433 (66%), Gaps = 13/433 (3%)

Query: 7 KSTRGKALLVIEAIFDQGYNTIALNQSLNKLAKDRALITRIVYGVTVSRKISLEWYL 66
 K+ R AL V+ IF AY NI+L++ L+ LS D+ +T +VYG VS+K LEWY+
 Sbjct: 3 KNARQTALDVLNDIFGNDAVANTISLDRNLRSDELSTVDKGFVTALVYGVVSKALLEWYL 62

Query: 67 AHYVDRDKLDKQVYVYLLMSLYQLTYLTKLEPAHIVNDVAVIAKRNKNGKGAELINAV 126
 +K K W LL++YQ++DK+P A V++AV IAK R + + P+NA+
 Sbjct: 63 TPLKKKEPK--PWAKMLLLTYYQLFMDKVPISAAVDEAVKIAK-RHDGQATANFINAV 119

-1087-

Query: 127 LRQFTSHPLPDMETIKRKNKYYSVKYSLPVLVKKLEQFGSDRSVAIMESLPVRSKASI 186
 LR F E K + KYS+P L+ K+ QFG R+ I+ESL S S+
 Sbjct: 120 LRNFMR-----EERNEEPKDWETKYSMPKLLDXMVRQFGGKRTGEILESIRKPSHVSL 174

5 Query: 187 RVTDPLKLEVAEALDARSLLSATGLTKASGHPAASDYFTNGDITIODESSQLVAPTN 246
 R DP E SLL+ T L SG+F+ ++ F G ITIQDE+SQLVAP L
 Sbjct: 175 RKIDP-----TWIAGTRPSLLTETALIDSGNFSITRETFQGRITIQDESSQLVAPQLE 229

10 Query: 247 IDGDDIILDACSAPGGKTSHTASGLTKGKIALDLYDHKLELVKRNARLGVADNITR 306
 ++G + +LDAC+APGGK++H+A YL TG + ALDLY+HKL+L+ +NA R VAD I T+K
 Sbjct: 230 LGTEERVLDCACAPGGKSTHMAQYLTGHTALDLYRHKLDDINQAQRQHVADKITQK 289

Query: 307 LDAREVHRHFEKDSFDKILVDAPCSGIGLIRKPKDKYNKESQGFNALQAIQLEILSSVC 366
 DA ++ +F + FD+ILVDAPCSGIGLIRKPKDI+Y KES F LQ IQLEIL+S
 Sbjct: 290 ADATMIYENPGPEKFDRIILVDAPCSGIGLIRKPKDIRYRKSSDFIDQLQLEILNSAS 349

15 Query: 367 QTLRKGGLIITYSTCTIFDEENRQVIRAFLOSHFNFBQVKLNHTQADIVKDGYLITPEQY 426
 ++L+K GI+ YSTCTIFDEEN V+ FL++HFNFBQV++ ++ ++K+L G+ ITPE Y
 Sbjct: 350 KSLKKGIMVYSTCTIFDEENFDVVEHFNFBQVLSNKRKPEVIRKGGCLFITPEY 409

20 Query: 427 QTDGFFIGQVRV 439
 TDGFFI + ++
 Sbjct: 410 HTDGFFIAKFKKI 422

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 305/440 (69%), Positives = 370/440 (83%)

Query: 1 MANDWKKSGARGIAIMTEKEVPDQKAGYSHIALNKSIAKSRSLSDKDRALVTEIYVGTVARKI 60
 +R+WKKS RG AL+ +E +FD+GAY+NIALN+ L LS KDRAL+TEIYVGT+VARKI
 30 Sbjct: 1 LADNWKSGTRGKALLVIEAIFDQGYTHIALNQQLSNKALSADKDRALLTEIYVGTVSRKI 60

Query: 61 TLEWYLSHFIVDRDKLELVVHLLLSLYQLLYLDNIPOHAIVNDVAITAKRGNKKGAE 120
 +LEWYL+H++ DRDKL+ WYV+L+LSLYQL YLD +P HAIVNDV IAKRGNKKGAE
 35 Sbjct: 61 SLEWYLAHYVKDRDKLDKVVYLLMLSLYQLTYLDKLPAAHVNDVAITAKRGNKKGAE 120

Query: 121 KLINAVLRVRSSETLPETASIKRKNKRYSVAYSMPVWLKGLIDQGETRALAIMESLP 180
 K +NA+LR+ +S LP++ +IKR+NK YSV YS+PWLKGL DQ+G R++AIMESLP
 Sbjct: 121 KPVNAILRQFTSHPLPDMETIKRKNKYYSVKYSLPVLVKKLEQFGSDRSVAIMESLP 180

40 Query: 181 RNKASLRVTDLSQKQKITETINVRDASHIAETALVADSGNFASTSPFQGLITIQDESSQL 240
 R+KAS+RVTD + + + E L+ S ++ T L SG+FA++ +F +G ITIQDESSQL
 Sbjct: 181 RSKAS+RVTDPLKLEVAEALDARSLLSATGLTKASGHPAASDYFTNGDITIODESSQL 240

45 Query: 241 VAPTLKVSQNDQVLDCASAPGGKTSHTASGLTKGKIALDLYDHKLELVKRNARLGVAD 300
 VAPTL + G+D +LDACSAPGGKTSHTASGLTKGKIALDLYDHKLELVKRNARLGVAD
 Sbjct: 241 VAPTLINIDGDDIILDACSAPGGKTSHTASGLTKGKIALDLYDHKLELVKRNARLGVAD 300

Query: 301 KIKTKKLDAKSAHEYFLESDPDKILVDAPCSGIGLIRKPKDKYNKANCOPRALQAIQLE 360
 I+T+KLDA + H +F +D+FDKILVDAPCSGIGLIRKPKDIKYNK +Q F ALQ IQL
 50 Sbjct: 301 NIETRKLDAREVHRHFEKDSFDKILVDAPCSGIGLIRKPKDKYNKANCOPRALQAIQLE 360

Query: 361 ILSSVCQTLRKGGLIITYSTCTIFDEENRQVIRAFLOSHFNFBQVKLNHTQADIVKDGYL 420
 ILSSVCQTLRKGGLIITYSTCTIF+EEN OVIE FL++HFNFBQV+L+HTQ DIVK G +
 Sbjct: 361 ILSSVCQTLRKGGLIITYSTCTIFDEENRQVIRAFLOSHFNFBQVKLNHTQADIVKDGYL 420

55 Query: 421 ISPEQVHTDGGFFIGQVRKL 440
 I+PEQY TDGFFIGQV+R+L
 Sbjct: 421 ITPEQYQTDGGFFIGQVRKL 440

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1088-

Example 986

A DNA sequence (GBSx1046) was identified in *S.agalactiae* <SEQ ID 3019> which encodes the amino acid sequence <SEQ ID 3020>. This protein is predicted to be pppL protein. Analysis of this protein sequence reveals the following:

```

5   Possible site: 45
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.5796 (Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:CAA10712 GB:AJ132604 pppL protein [Lactococcus lactis]
    Identities = 131/245 (53%), Positives = 177/245 (71%), Gaps = 4/245 (1%)

    Query: 1  NEISLLTDIGQRSSNNQDFINQFENKAGVPLIILADGMG3GHRAGNIASEMTVTDLGSDWA 60
      ME S+L+DIG +RS NQD++ + N+AG L +LADGMG3H+AGN+AS+TY DLQ N+
20   Sbjct: 1  NEYSILSDIGSKRSNTQDYVGTYNRAGYQLFLADGMG3GHRAGNVASKITVEDLGKIMS 60

    Query: 61  ETDFF--SELSEIRDMWILWSIETENRRIYELGQSDDYKMGTTIEAVALVGDNIIFAHVG 117
      ET F + + + W+ + EN I LG+ D+Y+MGTT+EA+ I G+ L+ AHVG
30   Sbjct: 61  ETTFFDAGTPEATLEIWLNRQVNRNENIASLGKLEDYQMGTTIEALVIKGNIVSARVG 120

    Query: 118 DSRIGIVRQGEYHLITSDHSLVNLVLKAGQLTETEEAASHPQRNIITQSIGQANFVEPDLG 177
      DSR ++R GE + +T+DHSLV ELV AGQ+TEEEA HP KNIIT+S+GQ N V+ D+
40   Sbjct: 121 DSRTYLMRQGLNKIITDHSLVQELVDAGQITETEEAASHVPHKNIITELGQTNVEQADIQ 180

    Query: 178 VHLEEGDYVAVNSDGLTNMISNADIAIVLQSK-TLDDKQDLITLANHGGGLDNITVA 236
      L- GD +++NSDGLTNM+S I VL +E TLD+K++ LI LAN GGLDNITV
50   Sbjct: 181 ALBLQAGDIILMNSDGLTNMVSITIMEVLEREDLTLNKSEALIRLANHGGGLDNITVV 240

    Query: 237 LVTVE 241
      L+ E
60   Sbjct: 241 LIKFE 245

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3021> which encodes the amino acid sequence <SEQ ID 3022>. Analysis of this protein sequence reveals the following:

```

40   Possible site: 43
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
        bacterial cytoplasm --- Certainty=0.5301 (Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
45   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 180/245 (73%), Positives = 220/245 (89%)

50   Query: 1  MEISLLTDIGQRSSNNQDFINQFENKAGVPLIILADGMG3GHRAGNIASEMTVTDLGSDWA 60
      M+ISL TDIGQ+RSNNQDFIN+P+NK G+ L+ILADGMG3GHRAGNIASEMTVDLG +W
      Sbjct: 1  MKISLKTIDIGQRSSNNQDFINQFENKAGVPLIILADGMG3GHRAGNIASEMTVTDLGREMV 60

    Query: 61  ETDFFSELSEIRDMWILWSIETENRRIYELGQSDDYKMGTTIEAVALVGDNIIFAHVGDSR 120
      +TDF+ELG+IRDMW +I++EN++IY+LGQS+D+KMGTT+EA+V +I+AH+GDSR
55   Sbjct: 61  KTDFTLSQIRDMWLFETIQSENQRIYDLGQSEDYKMGTTIEAVALVSSAIVAHIGDSR 120

    Query: 121 IGIVRQGEYHLITSDHSLVNLVLKAGQLTETEEAASHPQRNIITQSIGQANFVEPDLG 180
      IG+V G Y LLTSDHSLVNLVLKAGQ+TEEEAASHPQ+NIIITQSIOQA+PVEPDGV +
60   Sbjct: 121 IGLVHDGHYLLTSDHSLVNLVLKAGQITETEEAASHPQRNIITQSIGQANFVEPDGV 180

```

-1089-

Query: 181 LEEGDYLVVNSDGLTNMLSNADIAIVLTQEKTLDDKNQDLITLANHREGGLDNITVALVYV 240
 LE GDYLV+NSDGLTNM+SN +I T+L + +LD+KQ+I LAN RGLDNIT+ALV+
 Sbjct: 181 LEFGDYLVNSDGLTNMISNDEIVTLTGKSVSLDERKQEMIDLANRGLDNITVALVFN 240

5 Query: 241 ESEAV 245
 ESE V
 Sbjct: 241 ESEV 245

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 10 vaccines or diagnostics.

Example 987

A DNA sequence (GBSx1047) was identified in *S. agalactiae* <SEQ ID 3023> which encodes the amino acid sequence <SEQ ID 3024>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Ldkelihood = -10.03 Transmembrane 346 - 362 (340 - 372)

----- Final Results -----
 20 bacterial membrane --- Certainty=0.5012 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9539> which encodes amino acid sequence <SEQ ID 9540> was also identified.

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAAL0713 GB:AJ132604 hypothetical protein [Lactococcus lactis]
 Identities = 219/380 (57%), Positives = 284/380 (74%), Gaps = 8/380 (2%)

30 Query: 1 MIQIGKLFAGRYRILKSI GRGGMAVYLYARDLILNEEVAIKVLRITNYQTQDIARVFR 60
 MIQIGK+FA RYRI+K IGRGGMA+VY D L + +VAIKVLR+N++ D IA+ARFR
 Sbjct: 1 MIQIGKIFADRYRII KEIGRGGMANVYQGEDTFLGDRKVAIKVLRNSFNEDDIAIARFR 60

Query: 61 EARMAGELTHPNIVAIRDIGEEDGQFLVMEYVDGFLKCYIQNAPLGNNEVVRIMNEV 120
 EA AMAGEL+HPNIV I D+GE + QQ++VME+VDG LK+YI NAPL+N+E +I+ E+
 35 Sbjct: 61 EAFAMAGELSHPNIVGISDVGEFESQYITVMEFVDGNTLKQYINQAPLANDEAIEITSI 120

Query: 121 LSAHSLAHQKQIVHRDLKPCNLIITFKQTVKVTDFGLAVAFATSLAQTNMGLSVHLYS 180
 LSAH +AH GI+HRDLKPCN+L++ QTVKVTDFGLA A +ETSLAQTN+M GSVHLYS
 40 Sbjct: 121 LSAHMAHSHQIHRDLKPCNVLSSSQTVKVTDFGLAKALSETSLAQTNMFGSVHLYS 180

Query: 181 PEQARGSKATVQSDIYAMGIMLFEMIIGHIPYDGD SAVTIALQHPKPLPSILAENKEVP 240
 PEQARGS ATVQSDIYA+GI+LFE+LTG IP+DGD SAV IAL+HFO+ +PSI+ N VP
 Sbjct: 181 PEQARGSNATVQSDIYAGIILPELLTGQIPFDGDSAVAIALKHQPNIPISIIINAPFVP 240

45 Query: 241 QALENIVIKATAKILTRYKITYVNGRDLSTLSTHREPKLVN-DTESTIKLPKVT 299
 QALEN+VIKATAK +RY EM D++T+ S R E KLVN D +TK +P +
 Sbjct: 241 QALENVVTKATKADINRYADVENMTDVATSTSLDRRCBEKLVNKHDETKIMP--AN 298

50 Query: 300 TVSSLTTEQLLRNQKQAKTKITKIPDPSADNKTSSKKKASHRLSTINKLFFALCVGII 359
 ++ T+ L+ K+ EK +S++ + K+K K+S + G I+ L L V+G
 Sbjct: 299 LINPYDTKFLI--DKKTDQKAGSSSTTNNKNNKNNKSKK--GLISLVLVLVIGG 354

Query: 360 VFAYKILVSPPTIRVPDVSN 379
 FA+ + +PT ++VP+V+N
 55 Sbjct: 355 AFAMAV-STPTNVKVPNVIN 373

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3025> which encodes the amino acid sequence <SEQ ID 3026>. Analysis of this protein sequence reveals the following:

-1090-

Possible site: 56
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -8.60 Transmembrane 349 - 365 (340 - 370)

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.4439(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:
 >GP:CAA10713 GB:AJ132604 hypothetical protein [Lactococcus lactis]
 Identities = 209/378 (55%), Positives = 273/378 (71%), Gaps = 8/378 (2%)

15 Query: 1 MIQIGKLFAGRYRLKSIIRGGMADVYLANDLILNEDVAIKVLRNRYQTDQAVARFOR 60
 Sbjct: 1 MIQIGK+FA RYRI+K IGRGMA+VY D L+ VAIKVLN+N++ D +A+ARFQR 60
 MIQIGKIFADRYRIIIEIGRGMANVYQGEDFLGDRKVAIKVLRNFPENDDIAIARFOR 60

20 Query: 61 EARAMAELHNPINVAIRDIGEEGQQFLVMEYVDGADLKRYIQNHAPLSNNEVVRIMEEV 120
 Sbjct: 61 EARAMAELHNPINVI D+GE+ QQ++VME+VDG LK+YI +APL+N+E + I+ E+ 120
 EARAMAELHNPINVIIGISDVGEFSSQQYIVMEFVDGMLKQYINQNPANDAEALIEITEI 120

25 Query: 121 LSAMTIAHQKGIIVHRDLKPNILLTKGGVVKVTDPGIIVAFAPETSITQTINSLGASVHYLS 180
 Sbjct: 121 LSAM+AH GI+HRDLKPN+L++ G VKVTDPGIA A +ETSLTQIN+M GSVHYLS 180
 LSAMDMAHSGIIRHDLKPNVLSSSGTGVKVDGFIKALAKSETSLTQTINFGSVHYLS 180

30 Query: 181 PEQARGSKATIQSDIYAMGIMLFEMLTGHIPIYDGSVAITIALQHPKPLPSIIENHNVP 240
 Sbjct: 181 PEQARGS AT+QSDIYA+GI+LFE+LTG IP+DGSVA IAL+HPQ+ +PSII N VP 240
 PEQARGSKATVQSDIYAIIGILLPELLTGQIPFDGSAVAIALKHQPMISIIENHNVP 240

35 Query: 241 QALENVVIRATAKKLSORYGSTFEMSRDLTALSINRSRERKIIF+ENVESTKFLPKVAS 299
 Sbjct: 241 QALENVVI+ATAK +++RY EM D+T+ S+R E K++F ++ + TK +F 299
 QALENVVIKATAKDINNRYADVEEMTDVATSTSLDRGEEKLVFNKDHDETKIMFANLI 300

40 Query: 300 GPTASVKLSPPTTFLVLTQESRLDQINQTDALQPTTKKKSGRFLGTLFKLPSFPTQVA 359
 Sbjct: 301 NPYDTKPLIDKKTDD--DGEKAQSESSTENNNKNNKSKGGLISLVLL--VIGG 354
 P + L QE + + KK K G + +L + +G

Query: 360 LFTYLILKFTSVKVPNV 377
 F + + T PT+VKVPNV

40 Sbjct: 355 AFAMAVST-PTNVKVPNV 371

An alignment of the GAS and GBS proteins is shown below.

Identities = 390/643 (60%), Positives = 480/643 (73%), Gaps = 29/643 (4%)

45 Query: 1 MIQIGKLFAGRYRLKSIIRGGMADVYLANDLILNEDVAIKVLRNRYQTDQAVARFOR 60
 Sbjct: 1 MIQIGKLFAGRYRLKSIIRGGMADVYLA D LILNEDVAIKVLRNRYQTDQ+AVARFOR 60
 MIQIGKLFAGRYRLKSIIRGGMADVYLANDLILNEDVAIKVLRNRYQTDQAVARFOR 60

50 Query: 61 EARAMAELHNPINVAIRDIGEEGQQFLVMEYVDGDLK+YIQ++APLSNNEVVRIMEEV 120
 Sbjct: 61 EARAMAEL HNPINVAIRDIGEEGQQFLVMEYVDG DLK+YIQ++APLSNNEVVRIM EV 120
 EARAMAELHNPINVAIRDIGEEGQQFLVMEYVDGADLKRYIQNHAPLSNNEVVRIMEEV 120

55 Query: 121 LSAMTIAHQKGIIVHRDLKPNILLTKGGVVKVTDPGIIVAFAPETSITQTINSLGASVHYLS 180
 Sbjct: 121 LSAM+IAHQKGIIVHRDLKPNILLTK+G VKVTDPGIIVAFAPETSITQTINSLGASVHYLS 180
 LSAMTIAHQKGIIVHRDLKPNILLTKGGVVKVTDPGIIVAFAPETSITQTINSLGASVHYLS 180

60 Query: 181 PEQARGSKATVQSDIYAMGIMLFEMLTGHIPIYDGSVAITIALQHPKPLPSIIENHNVP 240
 Sbjct: 181 PEQARGSKAT+QSDIYAMGIMLFEMLTGHIPIYDGSVAITIALQHPKPLPSI+ EN +VP 240
 PEQARGSKATVQSDIYAMGIMLFEMLTGHIPIYDGSVAITIALQHPKPLPSIIENHNVP 240

65 Query: 241 QALENVVIRATAKKLSORYGSTFEMSRDLTALSINRSRERKIIF+ENVESTKFLPKVAS 299
 Sbjct: 241 QALENVVI+ATAKKL+DRY +T+EM RDL TALS R RE K++F + BSTK LKPV S 299
 QALENVVIRATAKKLSORYGSTFEMSRDLTALSINRSRERKIIFENVESTKFLPKVAS 300

65 Query: 300 -----TVSSLTTEQLLNQKQAKTTSKTIYDGSANDKTSKKKSHRLLGTIMKL 349
 T + LT E L Q T+ + P + KKK S R LGT+ K+

-1091-

Sbjct: 301 PTASVKLSPTTFFVLITQESRL--DQTNQTDALQPT-----KKKSGRFLGTLFKI 349

Query: 350 PFALCVVGLIIVFAYKILVSPITIRVVDVSNICVAQAKNTLNSGLKVGCAIRNIESDSVSE 409
 P+ +VG+ +F Y IL PT++VP+V+ ++ AK L + GLKVG IR IESD+V+E

5 Sbjct: 350 LFSFFIVGVALFTYLLITKPTSVKVPVNAQTSILKVAQQLYDVGKVGKIRQIESDITVAE 409

Query: 410 GLVVKTDPAAGSRREBAGKWNLYIATPNKSPITLGNVKEHYKDILKDL-QHGKVGKSLIK 468
 G VV+TDP AG ++R+G+ +LX+ NK F + NYK +Y++ + L + GV KS IK

10 Sbjct: 410 GNVVKTDPKAGTARQSSITLIVSIOGKQFDMEHYKGLDYORAMNSLIETVGVPSKIK 469

Query: 469 VKRKINNDYTTGTLAQSLPETSVPNPGNKLTITVAVNDPMIPQATIMVGEVITL 528
 ++R +N+Y T++QG G ENP+G K+IL+VAV+D + MP VT + + + TL

15 Sbjct: 470 IERIVTNEYPTNVISQPSAGDKPNPAGKSKITLGVAVSDTITMVMVTSYADAVNTL 529

Query: 529 TDLGLDADNLVFIQMGNGV---YQTVTPPSSSKIASQDPYTGGEVGLRRGDKVKLYLLG 585
 T LG+DA + Y + + + +P S + ++ Q FYG + L ++ LXL

Sbjct: 530 TALGIDASRIKAVPSSSSATGFVPIHSPSSKAIVSQSPYTGTSLSLSKGEISLYLP 589

Query: 586 SKTTNNSSTFIDSSASSSTGTTSDSVSSSTDASTDSSSTS 628
 +T ++SS+ SS SS ++ +DS + ++ S S +TS

20 Sbjct: 590 EETHSSSSS---SSTSSNSSSINDSTAFQSNTELSPETTS 629

SEQ ID 3024 (GBS297) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 43 (lane 6; MW 75kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 27 (lane 4; MW 100.2kDa) and in Figure 159 (lane 2-4; MW 100kDa). GBS297-GST was purified as shown in Figure 223, lane 3. GBS297-His was purified as shown in Figure 203, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

30 Example 988

A DNA sequence (GBSx1048) was identified in *S. agalactiae* <SEQ ID 3027> which encodes the amino acid sequence <SEQ ID 3028>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have an uncleavable N-term signal seq

35 INTEGRAL Likelihood = -7.91 Transmembrane 60 - 76 (50 - 90)
 INTEGRAL Likelihood = -7.43 Transmembrane 7 - 23 (3 - 25)
 INTEGRAL Likelihood = -5.68 Transmembrane 27 - 42 (24 - 46)

----- Final Results -----

40 bacterial membrane --- Certainty=0.4163 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:EAB03323 GAB035448 hypothetical protein [Staphylococcus aureus]
 Identities = 53/230 (23%), Positives = 104/230 (45%), Gaps = 14/230 (6%)

50 Query: 5 QFFLVAVAVLVLMGKMLSDQWTSPIIAL--ILLALRP--YNNDSRHFLITSLLL 61
 Q ++ A+++ I + P+ L L +L+ + Y + R LL+
 Sbjct: 9 QMLIIFTALMIANFYIYIFER-IGPLLVLGLCVLYVGYLYPEKIRGLAFWIGALLI 67

Query: 62 FLIFMLNPF-YIAAVFVAVLYVLINHPFSQVKKKRYALIQKMHQLDVKTTRNMGSTDQ 120
 + N Y II VF +L ++ + K K R + +K +W G

55 Sbjct: 68 APTLSKNYTIILQVPLLLIVRYLHKPKPKVATDEVMTSPSPK---QKWEGBQR 124

Query: 121 HESDFYAFEDINIRISGDTTDLTNVIVSQDNVYIIQKRVGDKVLVPLDVAADIS 180
 Y +ED+ I G IDLT ++N I++ + G +V+P++ + ++

-1092-

Sbjct: 125 TPVYVYVYEDVQIQHSGIGDLHDLTKAANKENNTIVVRHILGKVQVILPVNYNINLVA 184

Query: 181 SVYGSVQVYDFPEYDLRNESEIKLSQ--EREYLLKRVKLVVNTIAGKEV 228

+ YGS Y + + Y + N + I + + + + Y V + V + T G VEV

5 Sbjct: 185 APYGST-YVNEKSYKVENNNHIEEMMKPBNY---TVNIYVSTFIDGVEV 230

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3029> which encodes the amino acid sequence <SEQ ID 3030>. Analysis of this protein sequence reveals the following:

Possible site: 35
 10 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.92 Transmembrane 44 - 60 (36 - 64)
 INTEGRAL Likelihood = -8.76 Transmembrane 69 - 85 (66 - 105)
 INTEGRAL Likelihood = -8.70 Transmembrane 24 - 40 (20 - 42)
 15 INTEGRAL Likelihood = -6.64 Transmembrane 88 - 104 (85 - 105)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4970 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 20 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAB03323 GB:AB035448 hypothetical protein [Staphylococcus aureus]
 Identities = 41/187 (21%), Positives = 85/187 (44%), Gaps = 22/187 (11%)
 25 Query: 47 FILILVL--LILALRF-YNQDSRNFLITVSLFLPLMNPYIDMAVLGIVYPIKH 103
 F+L+L+L +L+ + + Y R + L + + + N Y I + + + + +
 Sbjct: 33 FLVILLGCVLVVGYLYPHKIRGLLAFWIGALLIAPTLLSNKTTIILFVLLLLIV-- 90
 30 Query: 104 PSQVKKIORFALIRPFKEKEIEVNT-----KHQWIGTANYSDFYCFDDIMIRISG 155
 R + + FK + K + K + N G Y + + D I G
 Sbjct: 91 -----RYLHKPKPKKVATDEVMTSFPIKQWPGQRTPVVYVGHEDVQIQHGIG 142
 35 Query: 156 NDTVDLTNIVITGMNINIIRKIPGNITILVPIDVTVILDVSSIVGSVDFPFCQQYDLRN 215
 + + DLT + N IV + R I G + + + P + + L V + + YGS + + Y + N
 Sbjct: 143 DLHIDLTKAANKENNTIVVRHILGKVQVILPVNYNINLVAAPYGST-YVNEKSYKVEN 201
 Query: 216 ESIKFKK 222
 + I + E
 40 Sbjct: 202 NNHIEE 208

An alignment of the GAS and GBS proteins is shown below.

Identities = 137/211 (64%), Positives = 175/211 (82%)
 45 Query: 1 MKKQFPFLLEAVVLVWGLMKILSDWTSFIFILALILALRFTNDSRNFLITLTSLL 60
 MKKQFPFLLE + + L MG + N IL + D + SF I L LILALRFTN DSR + NFLT SLL
 Sbjct: 18 MKKQFPFLLETCILLMGITMLTDLNLSLFFILILVILALALRFTNQDSRNFLITSLF 77
 50 Query: 61 LFLIFMLNPYIIAAVFAVLVYVLINHSQVKKKNRYALIQPKHQLDVKITRNQWIGTDQ 120
 LFLIFMLNPYII AV + + + Y INHSQVKKKNR + ALI + FK + + + V T + + QW + GT
 Sbjct: 78 LFLIFMLNPYIDMAVLGIVYIINHSQVKKKNR + ALI + FK + + + V T + + QW + GT 137
 Query: 121 HESDFYAFEDINIRISGTDITDLNVTVSGQDNVILIQYFGDITKVWPLDVAVKADIS 180
 +ESD + Y F + DINIRISG DT + DLNVTIV + G DN + I + K + FG + T + LVP + DV V D + S
 55 Sbjct: 138 YESDYCFDDINIRISGNTVDLNVITVGMNIIIRKIPGNITILVPIDVTVILDVSS 197
 Query: 181 SVYGSVQVYDFPEYDLRNESEIKLSQKREYLL 211
 S + YGSV + F + + YDLRNESEIK + + L
 Sbjct: 198 SIYGSVDFPFCQQYDLRNESEIKFSTENQSL 228
 60

SEQ ID 3028 (GBS66) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 6 (lane 4; MW 25kDa) and in Figure 7 (lane 2; MW 24.7kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 989

A DNA sequence (GBSx1049) was identified in *S.agalactiae* <SEQ ID 3031> which encodes the amino acid sequence <SEQ ID 3032>. This protein is predicted to be histidine kinase (narQ). Analysis of this protein sequence reveals the following:

```

Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
10  INTEGRAL    Likelihood = -11.41    Transmembrane    47 - 63 ( 40 - 72)
    INTEGRAL    Likelihood = -9.98    Transmembrane    9 - 25 ( 5 - 36)

----- Final Results -----
        bacterial membrane --- Certainty=0.5564 (Affirmative) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
15  bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP: CAB54570 GB: AJ006393 histidine kinase [Streptococcus pneumoniae]
Identities = 159/334 (47%), Positives = 239/334 (70%), Gaps = 5/334 (1%)

20  Query: 1  MKIGHYFLAFFYGVSVIIPALCFVIIDSLGVNL-VHLYQTSRLMWLIEQLIFSIFFLSLAVT 59
    MKK Y + . + +F + +L + +L+ + E+ +F + S++T
    Sbjct: 1  MKKQAVVIIALTSIFVFFFSHSLLEILDFWSIFLHVEKT---EKPFVLLLVFSMSMT 57

25  Query: 60  ILLLLTWFLDLDNSKQINENLRRIINANNQNSINVTDGTEISTNQRLSKQGNMTASLQS 119
    LL L W + + S R++ NL+R+L Q + D + + + LS K+NL+T +LQ
    Sbjct: 58  CLLALFWRGIEELSLRKNQANLKRLLAGQEVVQVAD-PDLASPKSLSGKLNLLTALQK 116

30  Query: 120 KENSRLKSQEVIVQERKRIARDLHDTVSQLFPAASMLSGIAQNSVQLVDQVQSLLA 179
    EN + + +EI+++ERKRIARDLHDTVSQLFPA M+LSGI+Q +LD + + +QL+ +
    Sbjct: 117 ARNQSLAQEEETIEKRIARDLHDTVSQLFPAAMILSGISQALKLDRKMTQLQS 176

35  Query: 180 VEENLQHAQNDLRILLHLRPVLEENKTLSEGFPMILKELTDKSDIEVVYHESITLPLKK 239
    V +L+ AQ DLR+LLHLRPVLE K+L EG + +LKEI DKSD+ V + + LPKK
    Sbjct: 177 VTAILETAQKDLRVLLHLRPVLEBQKSLIEGIQLLKELEDKSDLRVSLKQNMVTLPLKK 236

40  Query: 240 IEDNI FRIGQEFISNTLKHQSARLEVVYINQTNENILQKIMDNGIGFMDMSVYDLSYGLK 299
    IE+ +IFRI QE ISNTL+H+QAS L+VYL QT+ ELQLK++DNGIGF + S+ DLSYGL+
    Sbjct: 237 IEIEHIFRIQLISNTLRAHQASCLDVTLYQTVDLQKLVVDNGIGFQLGLDLDLSYGLR 296

45  Query: 300 NIEDRVEDLAGNQLLSQPKGVAMDIRLPLVNQ 333
    NI+ +RVED+AG +QLL+ P +G+A+DIR+PL+++
    Sbjct: 297 NIKERVEDMAGTVQLLTAQPKGLAVDIRIPLLEK 330

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2991> which encodes the amino acid sequence <SEQ ID 2992>. Analysis of this protein sequence reveals the following:

```

Possible site: 18
>>> Seems to have an uncleavable N-term signal seq
50  INTEGRAL    Likelihood = -14.22    Transmembrane    49 - 65 ( 42 - 70)
    INTEGRAL    Likelihood = -6.58    Transmembrane    8 - 24 ( 5 - 33)

----- Final Results -----
        bacterial membrane --- Certainty=0.6689 (Affirmative) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
55  bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 218/337 (64%), Positives = 276/337 (81%), Gaps = 3/337 (0%)

[illegible]

A related GBS gene <SEQ ID 8701> and protein <SEQ ID 8702> were also identified. Analysis of this protein sequence reveals the following:

```

30 Lipop: Possible site: -1  Crend: 4
    McG: Discrim Score: -14.69
    GVH: Signal Score (-7.5): -4.31
        Possible site: 19
    >>> Seems to have an uncleavable N-term signal seq
ALOM program count: 2 value: -11.41 threshold: 0.0
35 INTEGRAL Likelihood = -11.41 Transmembrane 47 - 63 ( 40 - 72)
    INTEGRAL Likelihood = -9.98 Transmembrane 9 - 25 ( 5 - 36)
    PERIPHERAL Likelihood = 3.61 146
    modified ALOM score: 2.78

40 *** Reasoning Step: 3

----- Final Results -----
        bacterial membrane --- Certainty=0.5564(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

52.5/77.6% over 288aa

Streptococcus pneumoniae

GP|5830526| histidine kinase Insert characterized

ORF00320(433 - 1302 of 1617)

GP|5830526|emb|CAB54570.1||AJ006393(43 - 331 of 331) histidine kinase {Streptococcus pneumoniae}

%Match = 28.6

%Identity = 52.4 %Similarity = 77.6

Matches = 152 Mismatches = 64 Conservative Sub.s = 73

252 282 312 342 372 402 432 462

QEEETP*NVSN*L*TLSELS*G*S*MKKHXYLFFAYFGYSVIFAICFVDSLGNLVHLYQTSRMVLSQILFISFPL

: : | : : : : | : : : :

MKQAYVVIALTSPFLPVPFSSHLSLILDEPMSIPLHDVKEFKVFLILWF

10 20 30 40 50

492 522 552 582 612 642 672 702

SLAVITILLILWFLLDDNSKQINENLRNLLNQSINVTDDGETISTNQLRSLKQNLMTASLOSKEKNSRIKVSQIVKO


```

5      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 10
```

SEQ ID 8702 (GBS31) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 15 (lane 8; MW 64kDa). It was also expressed as GBS31d in *E.coli* as a GST-
25 fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 151 (lane 8-10; MW 59kDa) and in Figure 187 (lane 8; MW 59kDa). GBS31d was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 151 (lane 11-13; MW 34kDa) and in Figure 182 (lane 11; MW 34kDa). Purified GBS31d-GST is shown in lane 3 of Figure 237.

Example 990

30 A DNA sequence (GBSx1050) was identified in *S.agalactiae* <SEQ ID 3033> which encodes the amino acid sequence <SEQ ID 3034>. Analysis of this protein sequence reveals the following:

Possible site: 31
>>> Seems to have no N-terminal signal sequence

```

35      ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2706(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

40 The protein has homology with the following sequences in the GENPEPT database.

>GF:CAB54571 GB:AJ006393 response regulator [Streptococcus pneumoniae]
Identities = 154/209 (73%). Positives = 184/209 (87%)

Query: 8 IKIVLVDDHEMVRLGLKSPINIQADVEVIGEASNGLEGIKKALELRPDVVMDLVMPED 67
45 +KI+LVDDHEMVRLGLK+ +IQ DVEV+GEASNG +GI ALELRPDV+VMD+VMPED+
Shift: 1 MKTILVDDHEMVRLGLKSPEDVQDVEVGEASNGSGDINALELRPDVTVMVMPED 60

Query: 68 GVEATLALLKDWPEAAILVLTSLDNEKIYPVIEAGAKGYMLKTSSAABILNAIRKVSRG 127
G++ATLA+LK+WPEA IL++TSLDNEKI PV++AGAKGYMLKTSSA E+LAA+ KV+ G
50 sbjct: 61 GIDATLALKKDWPEAKITVLTSLDNEKIMPVLDAGAKGYMLKTSSADELLHAAVSKVAAG 120

Query: 128 EQAIENEVDKKIKAHDKCPALHEGLTARERDILNLLAKGYDNQRIADRLFISLKTVKTHV 187
E AIE EV KK++ H LHE LTARERD+L L+AKGY+NQRIAD+LFISLKTVKTHV
Sbjct: 121 ELAIEQEVSKKVEYHRNHMELHEELTARERDVLQLIAKGYENQRIADDLFISLKTVKTHV 180

Query: 188 SNTLGKLNVDRTQAVVYAFQHLLVPQDD 216
SNIL KL V+DRTQA VYAFQHLLV Q++
Sbjct: 181 SNTLAKLEVSDRTQAAVYAFQHLLVGQEE 209

-1096-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2995> which encodes the amino acid sequence <SEQ ID 2996>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3094 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 175/212 (82%), Positives = 192/212 (90%)

Query: 5 MDKIKIVLVDDHEMVRVGLKSPINLQADVEVIGESNGLEGIKKALELRPDVVVMDLVMP 64
M KIK+LVDDHEMVR+GLKSPINLQAD+V+GEASNG EG+ AL L+PDV+VMDLVMP
Sbjct: 3 MSKIKIVLVDDHEMVRVGLKSPINLQADIDVVGEASNGREGVDLALALKPDVLVMDLVMP 62

Query: 65 EMDGVEATLALLKDWPEAAILVLTSLYLENEKIYPVIEAGAGYMLKTSSAAEILNAIRKV 124
E+ GVSEATL +LK W EA +LVLTSLYLENEKIYPVI+AGAGYMLKTSSAAEILNAIRKV
Sbjct: 63 ELGGVEATLEVLLKDWKEAKVLTSLYLENEKIYPVIDAGAGYMLKTSSAAEILNAIRKV 122

Query: 125 SRGEQALENEVDKKIKAHDKPALHBSGITARERDILNLLAKGYDNQRIADELFI SLKTVK 184
S+GE ALE EVDKKIKAHDK P LHE LTARE DIL+LLAKGYDNQ IADELFI SLKTVK
Sbjct: 123 SKGEALAIETVDKKIKAHDKHDLHBSGITAREYDILHLLAKGYDNQRIADELFI SLKTVK 182

Query: 185 TRVSNILGKLVNADRTQAVVYAFQHHLVPQDD 216
TRVSNIL KL V DRTQAVVYAF+HHLVPQDD
Sbjct: 183 TRVSNILAKLEVGDRTQAVVYAFRHHLVPQDD 214

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 991

A DNA sequence (GBSx1051) was identified in *S.agalactiae* <SEQ ID 3035> which encodes the amino acid sequence <SEQ ID 3036>. Analysis of this protein sequence reveals the following:

```

Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1688 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP: CAB06166 GB: Z94864 putative peptidyl-prolyl cis-trans isomerase
[Schizosaccharomyces pombe]
Identities = 81/174 (46%), Positives = 109/174 (62%), Gaps = 30/174 (17%)

Query: 288 IKTNHGDMIVKLFPDHAPKTVANFVIGLAKQGYDGLIFPHRIIPDFMIQGGDPITGTGNGGGE 347
++T+ G + ++L+ +HAPKT NF LAK+GYDGL+IFHR+IPDF+IQGGDPITGTG NG
Sbjct: 6 LQTSGLKILIELYTESHAPKTCQNFYTLAKGKYDGVIPHRVIPDFVIOGGDPITGTGRGCT 65

Query: 348 SIYGSFDEFSEELYNNV-RGALSMANAGPNTNSQPFIVQNTKIPYAKKELRGSGWPT 406
SIYG+ F+DE +L++ G LSMANAGPNTN SQFFI T P
Sbjct: 66 SIYGDKFDDEISDLHRTGAGILSMANAGPNTNSQFFI---TLAP----- 108

Query: 407 IAEYAGSGGTHPLDRRHISVFGVLQDSSFEVLDEIAAVETSGSQKPLEDVVIL 460
TP LD +H+FG+V S V + + T S D+P+E + I+
Sbjct: 109 -----TPWLQGKHTIPGRVV--SGLSVCKRMGLIRTDSSDRPIELPKII 150

```

-1097-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3037> which encodes the amino acid sequence <SEQ ID 3038>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2175 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

Identities = 381/464 (82%), Positives = 422/464 (90%)

```
Query: 1 MDAKTKYKAKKIKAVFFDIDTLRVKLTGYMPPSILKVFALKDKGIVGLASGRARYGV 60
MDAK KYKAKKIK VFFDIDTLRVKLTGYM SI +VFALK KGI+VGLASGRARYGV
Sbjct: 5 MDAKLKYKAKKIKAVFFDIDTLRVKLTGYMPSIQRFVFKALKAGILVGLASGRARYGV 64

Query: 61 PKEVQDLNADYCVKLNAGYVKDKDNIIFHRPIPAEYVEQYKKOMATVGIKYGLAGRHEA 120
P+EVQDL+ADYCVKLNAGYVKD K IIF PIPA+ V YKKOMAD +GI YG+AGRHEA
Sbjct: 65 PQEVQDLHADYCVKLNAGYVKDDAKTIIFQAPIPADVVVAYKKOMADMGIFYGMAGRHEA 124

Query: 121 VLSRDRLDVNDAIDIVSYDLEVNPDFNKEHDIYQMWTFSDEKDGSLHLEPIAEHLRLIRW 180
VLS R+D+++AID VY+ LEV PD+N+ HD+YQMWTFSDEKDG L LP LAEHLRL+RW
Sbjct: 125 VLSARNDMISNAIDNVYAQLEVCFDYNEHYHDVYQMWTFSDEKDGGLQPAELAEHLRLVRW 184

Query: 181 HDHSSDVLVKGTSKALGVSKVVEHLGLKPENILVFGDEINDELFDYAGLAVAMGVSHPE 240
HD+SSDVLVKGTSKALGVSKVV+HLGLKPENILVFGDEINDELFDYAG+++AMGVSHP
Sbjct: 185 HDHSSDVLVKGTSKALGVSKVVDHLGLKPENILVFGDEINDELFDYAGISLIMGVSHPL 244

Query: 241 AQKQADFITKKVEEDGILYALBSGLI+KELFPQ+D+N+G+P A IKTNGDMT+LVLF 300
Q+KADFITKKVEEDGILYALBSGLI+KEL FPQ+D+N+G+P A IKTNGDMT+ LVLF
Sbjct: 245 LQEKADFITKKVEEDGILYALBSGLIDKELQFPQLDLPNHKGPKATIKTNGDMT+LVLF 304

Query: 301 PDHAPKTVANFGLAKGYDGIIPHRIIPDFMIQGGDPTGTGMCQGSIGESFEDEFSS 360
PDHAPKTVANF+GLAK+GYDGIIPHRIIP+FMQGGDPTGTG G+SIGESFEDEFSS+
Sbjct: 305 PDHAPKTVANFGLAKEGYDGIIPHRIIPDFMIQGGDPTGTGMCQGSIGESFEDEFSD 364

Query: 361 ELYNVRGALSMANAGPNTNGSOFFIVQNTKIPYAKKELRGGWFTPIAEIYAGCGGTGPHL 420
ELYN+RGALSMANAGPNTNGSOFFIVQN+KIPYAKKELRGGWP PIA YA GGTGPHL
Sbjct: 365 ELYNLRGALSMANAGPNTNGSOFFIVQNSKIPYAKKELRGGWNPAPIASYANGGTGPHL 424

Query: 421 DRRHSVFGQLVDQSSFEVLDEIAAVETGSQDKPLEDVILTIEV 464
DRRH+VFGQLVD++SF+VLD IA VETG+QDKP EDV+I TIEV
Sbjct: 425 DRRHTVFGQLVDQSSFEVLDEIAAVETGQDKPKEDVIITIEV 468
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 992

A DNA sequence (GBSx1052) was identified in *S.agalactiae* <SEQ ID 3039> which encodes the amino acid sequence <SEQ ID 3040>. This protein is predicted to be ribosomal protein S1 (rpsA). Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3126 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

-1098-

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB07066 GB:AP001518 polycyribonucleotide nucleotidyltransferase
(general stress protein 13) [Bacillus halodurans]
Identities = 46/120 (38%), Positives = 71/120 (58%), Gaps = 11/120 (9%)
Query: 8 KIGDKLKQTVTIGIRPYGAFVSLDGRTGLIHISEIKTGYIDNIYDVLVSGDEVYVQVIDV 67
++G ++G VTGI+P+GAFV++D + GL+HISE+ G++ +I DVLVSGDEV V+++ V
Sbjct: 5 EVGSGIVSGKVIGIKPFGAFVAIDDDQKGLWHISEVANGFVKDINDVLVSGDEVVKILSV 64
Query: 68 DEFTQKASLSRLTLEERHHIQH-----RIIRFNNRLKIGFKPLENLPSWVEE 116
DE + K SLG+R +E R GF LE+ L W+++
Sbjct: 65 DEESGKISLSIRATQAPERPARAKPKFPAGGGGRKPKQGQGGQGGNTLEDKLEMLKQ 124

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3041> which encodes the amino acid sequence <SEQ ID 3042>. Analysis of this protein sequence reveals the following:

Possible site: 58
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
bacterial cytoplasm --- Certainty=0.1832 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 78/115 (67%), Positives = 100/115 (86%)
Query: 7 MKIGDKLKQTVTIGIRPYGAFVSLDGRTGLIHISEIKTGYIDNIYDVLVSGDEVYVQVID 66
MKIGDKL GT+TGT+PYGAFV+LE+G TGLIHISEIKTG+ID+I +L++G++V VQVID
Sbjct: 1 MKIGDKLHGTITGINKPYGAFVLENGTTGLIHISEIKTGFIDIDQLLAIGNQVLVQVID 60
Query: 67 VDEFTQKASLSRLTLEERHHIQHRRFNNRLKIGFKPLENLPSWVEEGLAYL 121
+DE+++K SLG+RTL EE+ H HRR+SN+R KIGF+PLEE LP W+EE L +L
Sbjct: 61 IDEYSKKPSELMTLAEEKQHPFRRHRYNSNRHKIGFRPLEEQLPQNIIEESLQFL 115

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 993

A DNA sequence (GBSx1053) was identified in *S.agalactiae* <SEQ ID 3043> which encodes the amino acid sequence <SEQ ID 3044>. This protein is predicted to be pyruvate formate-lyase 2 activating enzyme (pflA). Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
bacterial cytoplasm --- Certainty=0.2889 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC76934 GB:AB000469 probable pyruvate formate lyase activating
enzyme 2 [Escherichia coli K12]
Identities = 90/251 (35%), Positives = 142/251 (55%), Gaps = 16/251 (6%)
Query: 8 VFNIQHPSTHDGPIRTITVFLKGCPLCFWCANPESQRMVPEPTR----- 52
+FNIQ +S++DG GIRT VF KGCP CFWCANPES +T+R
Sbjct: 24 IFNIQRYSLNDGEGIRITVVFVFKGCPHLCFWCANPESISGKIQTIVRRACKLHCACKLRDA 83

-1099-

Query: 53 -DAITNESVIVGEEKSVDDIIEEVLKIDIDFYEESSGGGITLSGGEIFAQFEFAKAILKRAK 111
 + + +G+S+D+ EV+KD P+ SGGG+TSGGE+ Q EFA L+R +
 Sbjct: 84 DECPGSAFERIGRDISLDALEKEVKKDDIFPRTSGGGVTLSSGEVLMQARFATRFQLRLR 143

5 Query: 112 SLGIHTAETTTATREHQFIDLIQYVDIYFDLKHYNLSLKHQKTMVKNASIIKNIHYAF 171
 G+ AIET + + L+ D+ DLK ++ ++ + ++N+
 Sbjct: 144 LMGVSCAIEAGDAPASFKLLPLAFLCDEVLFDLKMIDATQARDVVVKNLPRVLLENLELLV 203

10 Query: 172 ANGKTIVLRIPVPIPNFNDSLDKEPAFLDRLDIRQVQLLPHFHQGNKYQLLNQRYEM 231
 + G ++ R+P+IP F S E+ ++ + L+IRQ+ LLPHQ+G+ KY+LL + + M
 Sbjct: 204 SBGVNVIPLRLPIPGFILSRNNQQALDVLIPLNIRQHILLPHFHQGEFKYRLGLGKTWSM 263

Query: 232 EEIALHPEDL 242
 +E+ A D+
 15 Sbjct: 264 KEVPAPSSADV 274

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3045> which encodes the amino acid sequence <SEQ ID 3046>. Analysis of this protein sequence reveals the following:

Possible site: 58
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2209 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 187/255 (73%), Positives = 220/255 (85%)

30 Query: 4 EKGIVFNHQHFSIHGPGGIRTTIVFLKGCPLRCIPWCANPESQKQVFMTRDAITNESVIVG 63
 ++GIVFNHQHFSIHGPGGIRTTIVFLKGCPLRCIPWCANPESQ+ PE M + + IVG
 Sbjct: 3 DRGIVFNHQHFSIHGPGGIRTTIVFLKGCPLRCIPWCANPESQQAQECMLTSDGLNTRIVG 62

35 Query: 64 EEKSVDDIIEEVLKIDIDFYEESSGGGITLSGGEIFAQFEFAKAILKRAKSLGIHTAETTTA 123
 EEK+VD++IEEVLKID+DFYEESSGGG+TSGGEIFAQF+FA A+LK AK+ G+HTAETTTA
 Sbjct: 63 EERKTVDVEIEEVLKIDIDFYEESSGGGNTLSGGEIFAQDFDALALLKAAKAGLHTAETTTA 122

Query: 124 YTRHEQFIDLIQYVDIYFDLKHYNLSLKHQKTMVKNASIIKNIHYAFANGKTIVLRIPV 183
 + +HEQF+ L+ YVDFIYFDLKHYN L+HQ+ T V+N IIKNIHYAF GK IVLRIPV
 40 Sbjct: 123 FAKHEQFVTLVDYVDIYFDLKHYNLQHEKVTGVRNDLIKNIHYAFQAGKEIVLRIPV 182

Query: 184 IPNFNDSLDKEPAFLDRLDIRQVQLLPHFHQGNKYQLLNQRYEMEEIALHPEDLL 243
 IP FNDSL+DA+ F+ LF+L+LI QVQLLPHFHQF+G+NKY+LL R+YEM E+ A HPEDL
 45 Sbjct: 183 IPQFNDSLDCAKASELFNQLRIDQVQLLPHFHQFGENKYKLLREYEMAEVKAHPEDLA 242

Query: 244 DYQAIFSKYNIHCYF 258
 DYQA+F +NIHCYF
 Sbjct: 243 DYQAVFLNHNHCYF 257

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 994

A DNA sequence (GBSx1054) was identified in *S.agalactiae* <SEQ ID 3047> which encodes the amino acid sequence <SEQ ID 3048>. Analysis of this protein sequence reveals the following:

55 Possible site: 57
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1762 (Affirmative) < succ>
 60 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-1100-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9299> which encodes amino acid sequence <SEQ ID 9300> was also identified.

- 5 The protein has homology with the following sequences in the GENPEPT database.

>GP:AC74366 GB:AB000226 putative DBOR-type transcriptional
regulator [Escherichia coli K12]
Identities = 74/177 (41%), Positives = 113/177 (63%), Gaps = 1/177 (0%)

- 10 Query: 2 MNRLNIIISLVSQYQKIDVNTLSSELLQVSKVTIRKDLKLEKGLLREHGAVLNSGDD 61
+R + I + +V ++ V L++ VS+VTIR+DL+ LE L R HG+AV DD+
Sbjct: 3 SQQTILQVVIDQQQVSVDLAKATGVSEVTIRQDLNLTLEKLSYLRAHGFVAVSLSDCV 62
- 15 Query: 62 NVALSPNHHKTKKEIAALANMVSDNDTILIESGSTCALLAENICQTKRNVITILNSCFIA 121
R+ N+ K+E+A AA++V +TI IE+GS+ ALA + + K+NVTI+T S +IA
Sbjct: 63 ETRQSNVTILKRELAEFAASLVQPGETIFIEGSSNALLARTLGEQKQNVITITVSYIA 122
- Query: 122 NYLREYDSQCQIVLLGGEYQSSQVTVGPELLKRMISLFHVSLAFVGTGDFDPKTRIYG 178
+ L++ C++ILGG YQ S+ VGPL ++ I H S AF+G DG+ P+T G
20 Sbjct: 123 HLLKD-APCEVILLGGVYQKXSESNGVPLTRQCIQQVHFSKAFIGIDGWQFETGFTG 178

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3049> which encodes the amino acid sequence <SEQ ID 3050>. Analysis of this protein sequence reveals the following:

- Possible site: 58
25 >>> Seems to have no N-terminal signal sequence
- Final Results -----
bacterial cytoplasm --- Certainty=0.2888 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- Identities = 131/171 (76%), Positives = 150/171 (87%)
- 35 Query: 1 MNRLNIIISLVSQYQKIDVNTLSSELLQVSKVTIRKDLKLEKGLLREHGAVLNSGDD 60
MNRL I I LVSQ +KIDVN+LSE L VSKVTIRKDLKLE KGLL REHGAVLNSGDD
Sbjct: 2 MNRLERIIQLVQKKKIDVNSLSEQLDVSKVTIRKDLKLEKGLLREHGAVLNSGDD 61
- 40 Query: 61 LNVLSPNHHKTKKEIAALANMVSDNDTILIESGSTCALLAENICQTKRNVITILNSCFI 120
LNVL S+N+ K- IA AA +V DNDTI+IESGSTCALLAE +CQTKRN+ ++TNSCFI
Sbjct: 62 LNVLSYNYNIKKRIAEKAEVLQDNDTIMIESGSTCALLAEVLQCTKENIKVITNSCFI 121
- Query: 121 NYLREYDSQCQIVLLGGEYQSSQVTVGPELLKRMISLFHVSLAFVGTGDFD 171
ANY+R+Y SCQI+LLGG YQ +S+VTVGPELLK+MISLFHV+ FVGTGDF+
45 Sbjct: 122 NYIRQYSSCQIILLGGYQPNSEVTVGPELLKRMISLFHVHVFVGTGDFN 172

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 995

- 50 A DNA sequence (GBSx1055) was identified in *S.agalactiae* <SEQ ID 3051> which encodes the amino acid sequence <SEQ ID 3052>. Analysis of this protein sequence reveals the following:

Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1672 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-1101-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AG04879 GB:AB004578 probable transcriptional regulator
[Pseudomonas aeruginosa]
Identities = 20/70 (28%), Positives = 40/70 (56%)

Query: 6 GPMGRDLMRSEVAQEMANADEVILLTDSKFNQATLVBLPLSTVSVQVITDKHENSEIA 65
G M + +E+A+ M A ++ ++ DSSK + AL + PLS +++++ D+ P E+

10 Sbjct: 179 GAMDPSIEAEIARAMTAQARQLTVIADSSKIGRRALFQVFLGRINRLVDRKPTGELW 238

Query: 66 NLFQEAERITI 75
Q+A + +

15 Sbjct: 239 EALQARVEV 248

There is also homology to SEQ ID 3050.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 996

- 20 A DNA sequence (GBSx1056) was identified in *S.agalactiae* <SEQ ID 3053> which encodes the amino acid sequence <SEQ ID 3054>. This protein is predicted to be transcriptional regulator. Analysis of this protein sequence reveals the following:

Possible site: 46
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0904 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30

A related GBS nucleic acid sequence <SEQ ID 9541> which encodes amino acid sequence <SEQ ID 9542> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:BA004499 GB:AP001509 transcriptional regulator [Bacillus halodurans]
Identities = 98/309 (31%), Positives = 178/309 (56%), Gaps = 1/309 (0%)

Query: 6 ERQKLLAKVAYLYYMBGKSQSEIANELGIYRTTISRMLAKAREEGLVRIEISDPNPEIFQ 65
E ++L+ KVA LYY EG +Q++A ++G+ R IS++L KA+E+G+V I I D N +

40 Sbjct: 5 EERRLLIVKASLYYFEGTQQAQVAKKIGVSRPVISKILNKAKKEQXIVYIKDKENHTIVR 64

Query: 66 LESYFKSKYHLKDIEIVSRKDSITSEIKDLAHVAAAMIRKKIKENDKVGIAWGRTLSK 125
LE + KYHLK+ +V + I++ + + + K IK D +GI+WG T+S

45 Sbjct: 65 LEQRLKCKYHLKRAIVVPT-SGLTQDMIKRAIGKATSYVSKNIKMGDSIGISWGTTVSS 123

Query: 126 VVEAMRPHPVSVQSVFVPLAGGSPHINAKYHVNTLVYEMRRFPQSGCTFINATLVQENANL 185
V+ ++ +PL GG H N L YE+++ C++ A + E L

50 Sbjct: 124 FVQETPYEQHRELKVIPLVGGWIRKPFVELASNLAYELAKOMNCRCSYLYAPAMVEAKEL 183

Query: 186 AKGILTSKYFBLMDNWEKLDVAIVGVGKPKESNBQQWLDLLNQDDPQCLDERAAVGRIT 245
++ S+ +++ ++A+VG+G K + + + L ++D L + AVG++

55 Sbjct: 184 KERLIQSEDIASVLEEGRNVMKAVGTGSPFGKSTMVNTLKERDITATLKIKI GAUGLMS 243

Query: 246 CRFFNHSGDVPVQHLAKRTTIGITLEQLQKVPNRIVAHGNYKAAALIAVLKGYINHLVT 305
RF++ G P++ L + IGI L++L++P I V+ G+K ++ A LK GY++ LVT

Sbjct: 244 SRFYDALGQPIDHPLNELVIGIDLELAKRIPITIVIGVSEGAHKVDSVEAALKGYIIDLVT 303

Query: 306 DFSTALNLT 314

-1102-

D STA +++
 Sbjct: 304 DDSTAQLI 312

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3055> which encodes the amino acid sequence <SEQ ID 3056>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2123 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- 15 Identities = 165/324 (50%), Positives = 238/324 (72%), Gaps = 1/324 (0%)
 Query: 3 MKLERQKLLAKVAVLYLYMEGKSGSEIANELGIYRTTISRMLAKAREBGLVRIETSDPNE 62
 MK ER++LLAKVAVL-Y++GKSQ+ I+ E+ TYRKT+ EMLAKA+EEG+VRIETI+D++ +
 Sbjct: 1 MKERRRLLAKVAVLYHYVQKSGQLISKEMNIYKTTVCRLAKAREBGLVRIETIADYDAD 60
 20 Query: 63 IPQLESYFYSKYHLEDIEIVSRKDSITSEIEKDLARVAAMIRKICKENDKVGILANGRT 122
 +F LE Y + +Y L+ +++V ++ + + + +A AA + R +K+ DK+G++WG T
 Sbjct: 61 LFALEHYVRQQYGLKLDIVPAQVEDTFMDTLTNVAKTAAEVFRHVVDKGDKIGISWGAT 120
 25 Query: 123 LSKVVEAMRPHFVSQVSFVPLAGGSPSHINATYVNTLVYEMSRFPQGSCTFINATLVQEN 182
 LS +++ + P + V PLAGGSPSHINA+YRVNTLVY ++R F G+ P+NA ++QE+
 Sbjct: 121 LSCIMDELNPKMKDVFIYPLAGGSPSHINAKYKVTNLVYRLARIPHGNSAFPMNAVIG 180
 30 Query: 183 ANLAKGILTSKYFEGIMDNWEKLDVAIVGVGGKPKSNEQ-QWLIDLMLQDDPQCLDEEPAV 241
 +LAKGIL SKYF ++ +W++LD+A+VG+G+P S RQ QM DLL D L E AV
 Sbjct: 181 KHLAKGILQSKYFNDILTSWDQLDLALVIGIGEPNSLEQSQMRDLTSSDHQLKYEKAV 240
 35 Query: 242 GEITCRFFNHSQDPVNVQHLAKRTIGITILEQLQKVPNRIVAHGNYKAALLAVLKKGYIN 301
 GE+ CRFF+ +G PV L RTIGI+LEQL+VP +AVA G +KA A+LA LK G+IN
 Sbjct: 241 GEVCCRFFDQAGQPVYTGILQRTIGISLEQLERVPKTNVAVATGKHAKAIALAALVAGPIN 300
 Query: 302 HLWTFSTALNILLRLKDKTFVDTI 325
 +LWTD T L +L LD+D ++ +
 Sbjct: 301 YLWTDKETMLAVLALDEDIDLNV 324
 40

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 997

- 5 A DNA sequence (GBSx1057) was identified in *S.agalactiae* <SEQ ID 3057> which encodes the amino acid sequence <SEQ ID 3058>. This protein is predicted to be PTS enzyme III cel (celC). Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have a cleavable N-term signal seq.

- 50 ----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 55 A related GBS nucleic acid sequence <SEQ ID 9543> which encodes amino acid sequence <SEQ ID 9544> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1103-

>GP:AAA23551 GB:M93570 PTS enzyme III cel [Escherichia coli]
Identities = 42/102 (41%), Positives = 70/102 (68%)

Query: 4 EIIVADQIIIMGLIINAGDAKQHIYQALKAKGKNFARSKIETIELADSALLRAHNLQTQFL 63
E+ +++MGLI+N+G A+ Y ALK AK+G+FA +K ++ + AL EAH +QT+ +
Sbjct: 13 EVERLEEEVVMGLIINSGARSLAYAAKQAKQGDFAAKAMMDQSRMAINEAHLVQTKLI 72

Query: 64 AQEAGGTRTDSALFIHQDHLMTSITRINLIKRIIDLRQEL 105
+AG + +S + +H+QDHLMTS+ LI E+I+L ++L
Sbjct: 73 EGDAGEGKMKVSLVHVAQDHLMTSMRLARELITELTEHKEG 114

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3059> which encodes the amino acid sequence <SEQ ID 3060>. Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC74806 GB:AE000268 PEP-dependent phosphotransferase enzyme III
for cellobiose, arbutin, and salicin [Escherichia coli]
Identities = 39/97 (40%), Positives = 66/97 (67%)

Query: 7 DQIIIMGLIINAGDAKQHIYQALKKAKEDDYATSEKEMALDALLRAHNLQTQFLAQEAS 66
+++MGLI+N+G A+ Y ALK AK+D+A ++ M + AL EAH +QT+ + +A
Sbjct: 18 BEVVMGLIINSGARSLAYAAKQAKQGDFAAKAMMDQSRMAINEAHLVQTKLIEGDAG 77

Query: 67 GNKSEITLFPVHSQDHLMTTITTEINLIKEIIDLRKEL 103
K +++ + VH+QDHLMT++ LI E+I+L ++L
Sbjct: 78 EGRKVSLSLVHVAQDHLMTSMRLARELITELTEHKEG 114

An alignment of the GAS and GBS proteins is shown below.

Identities = 81/103 (78%), Positives = 94/103 (90%)

Query: 3 MEIIVADQIIIMGLIINAGDAKQHIYQALKAKGKNFARSKIETIELADSALLRAHNLQTQF 62
M++IV DQIIIMGLIINAGDAKQHIYQALK AKS ++A S+ E+ LAD ALLEAHNLQTQF
Sbjct: 1 MQVIVPDQIIIMGLIINAGDAKQHIYQALKAKKEDDYATSEKEMALDALLRAHNLQTQF 60

Query: 63 LAQEAGGTRTDSALFIHQDHLMTSITRINLIKRIIDLRQEL 105
LAQEA G +++I+ALF+HSQDHLMT+ITEINLIKRIIDLR+EL
Sbjct: 61 LAQEAGGNKSEITLFPVHSQDHLMTTITTEINLIKEIIDLRKEL 103

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 998

A DNA sequence (GBSx1058) was identified in *S.galactiae* <SEQ ID 3061> which encodes the amino acid sequence <SEQ ID 3062>. This protein is predicted to be PTS system, cellobiose-specific IIB component (celA). Analysis of this protein sequence reveals the following:

Possible site: 24
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- >GP:AAF94440 GB:AE004207 PTS system, cellobiose-specific IIB
component [Vibrio cholerae]
Identities = 46/100 (46%), Positives = 62/100 (62%)
- Query: 1 MIKIGLFCAGFSTGMLVNMKIAADKEGIRAHIRAYSCGKIADYAKDLVDALLGPQVSY 60
N KI L C+AG ST MLV N+ AA+ +GIE I+A S + ++ DV LLGPQV +
Sbjct: 1 MKKILLCCSAGMSTSMVLVNMKQAESKGIECKIDALSVAFAEIRAIQETDVCLLGPQVRF 60
- Query: 61 TLDSKSIKDEYGVPIAVIPMDYGMGLDGKVLKLSLL 100
L++ + DEYG IA I YGM+ G +VL+ AL L+
Sbjct: 61 QLEELRKTADEYGNIAAISPOAYGMKMGDEVLQALDLI 100
- 15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3063> which encodes the amino acid sequence <SEQ ID 3064>. Analysis of this protein sequence reveals the following:
- Possible site: 31
>>> Seems to have a cleavable N-term signal seq.

- 20 ----- Final Results -----
bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 25 The protein has homology with the following sequences in the databases:

- >GP:AAF94440 GB:AE004207 PTS system, cellobiose-specific IIB
component [Vibrio cholerae]
Identities = 43/100 (43%), Positives = 58/100 (58%)
- 30 Query: 8 MIKIGLFCAGFSTGMLVNMKIAAEKKGIDCQIRAYACQKIADYAPLLVDALLGPQVAY 67
M KI L C+AG ST MLV N+ AAE KGI+C+I+A + + DV LLGPQV +
Sbjct: 1 MKKILLCCSAGMSTSMVLVNMKQAESKGIECKIDALSVAFAEIRAIQETDVCLLGPQVRF 60
- 35 Query: 68 TLDSKSAICKNDIPIAVIPMDYGMGLDGKVLKLSLL 107
L++ + IA I YGM+ G+VL AL L+
Sbjct: 61 QLEELRKTADEYGNIAAISPOAYGMKMGDEVLQALDLI 100

An alignment of the GAS and GBS proteins is shown below.

- Identities = 79/101 (78%), Positives = 92/101 (90%)
- 40 Query: 1 MIKIGLFCAGFSTGMLVNMKIAADKEGIRAHIRAYSCGKIADYAKDLVDALLGPQVSY 60
MIKIGLFCAGFSTGMLVNMK+AA+K+GI+ IBAY+QK+ADYA LDVALLGPQV+Y
Sbjct: 8 MIKIGLFCAGFSTGMLVNMKVAEKKGIDCQIRAYACQKGLADYAPLLVDALLGPQVAY 67
- 45 Query: 61 TLDSKSIKDEYGVPIAVIPMDYGMGLDGKVLKLSLL 101
TLDSK++IC + +PIAVIPMDYGMGLDG KVL LALSL++
Sbjct: 68 TLDSKSAICKNDIPIAVIPMDYGMGLDGKVLKLSLL 108

- SEQ ID 3062 (GBS180) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 39 (lane 4; MW 12.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 2; MW 37.6kDa).

The GBS180-GST fusion product was purified (Figure 204, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 298), which confirmed that the protein is immunoreactive on GBS bacteria.

- 55 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 999

A DNA sequence (GBSx1059) was identified in *S. agalactiae* <SEQ ID 3065> which encodes the amino acid sequence <SEQ ID 3066>. This protein is predicted to be pts system, cellobiose-specific iic component (celB). Analysis of this protein sequence reveals the following:

```

5      Possible site: 40
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -11.68      Transmembrane 346 - 362 ( 334 - 374)
      INTEGRAL      Likelihood = -9.77       Transmembrane 182 - 198 ( 178 - 205)
10     INTEGRAL      Likelihood = -8.65       Transmembrane 29 - 45 ( 27 - 50)
      INTEGRAL      Likelihood = -6.53       Transmembrane 140 - 156 ( 134 - 161)
      INTEGRAL      Likelihood = -4.78       Transmembrane 292 - 308 ( 289 - 312)
      INTEGRAL      Likelihood = -4.41       Transmembrane 397 - 413 ( 395 - 416)
      INTEGRAL      Likelihood = -2.97       Transmembrane 77 - 93 ( 72 - 93)
15     INTEGRAL      Likelihood = -2.97       Transmembrane 228 - 244 ( 222 - 246)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.5670 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
20     bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AA17390 GB:U07818 cellobiose phosphotransferase enzyme II''
[Bacillus stearothermophilus]
25     Identities = 160/415 (38%), Positives = 251/415 (59%), Gaps = 13/415 (3%)

Query: 15  KFNVMRGIIALKDGMILAILPLTVVSGSLFLILQQLPFGKLNQAINVFGPEWTEPFMVQVYS 74
      K   R + A++DG++ +PL ++GSLFLI+G LP G N++ A+ FG W + +
Sbjct: 18  KIRQRHLQAIRDGIILSMFLIILIGSLFVLIVGLFPIPGYNEWNAKWFGEHRLDKLLYPVG 77

30     Query: 75  GTFAIMGLISCPALAYAYAKNSVEPLPAQVLSLSSFFILMKSSYPVKGEA-----IA 128
      TF IM+L F +AY A+ V+ L AG +SL++F +L + P E ++
Sbjct: 78  ATFDINALVVSFGVAYRLAEKYKVDALSAGALSIAAF-LIATPYQVPTPEGAKRTIMVS 136

35     Query: 129 DAISKVMFGGQIGIIGIILVVGAIYIWFQIHVIRIMPEQVQPAIAKQFEAMIPAFVI 160
      I W G +G+ A+I+ +V IY IQ +IVIK+P+ VP A++A F A+IP +
Sbjct: 137 GGIPVQVWGSKGLFVAMILAIIVSTRIYRKIQKNIVIKLPGDVPPEAVARSFVALIPGAHV 196

Query: 189  FLLSMIVYLIKAVTTGGTFEMDIYIQVPLQGLTGSLYGAIGIAFFISFLMWFGVHQQS 248
      ++ + LI ++T +F ++ ++ PL L GS++GAI + LW G+HG +
40     Sbjct: 197 LVVVVARLILEMTFFESPHNIVSVLANKPLSVLQSGVFGAIVAVLVQLMSGLHGA 256

Query: 249  VVNGIVTALLLSNLDANKSLAAN-RITLNGAHIVTQCFDLSFLILSGSQTIFGLVIAM 307
      +V G++ + LS +D N++ N L N ++TQGF D ++ + GSG T L + M
Sbjct: 257 IVGGVMGPWILSLMDENRMVFGQNPMAELN---VITQGFDLWIVYIGSGSGATLALATM 313

45     Query: 308  LFAAKSKQYKALGKVAAPFAIFNVNEPIVGFPIVMVPMFLPFLIVFVLAALIVYGAIA 367
      +F A+S+Q K+LG++A P IIN+NEPI PG PIVMNP++ +PFLVPV+ ++ Y A+A
Sbjct: 314 MPRARSRLQKLSGRALIAFGPIFNINEPIVPGMPTVMNPLLIIPFLIVPVVIVVVSAYAMA 373

50     Query: 368  VGFMPQPPSGVILFWSTPAISGFMVGGWQ--GALVQIVILAISTAVVFFPFKIQD 420
      G + SGV +PW+TP +ISG++ G + G++QIV I+ A+Y+PFF I D
Sbjct: 374 TGLVAKPSGVAVPWVTPVIVSGYLATGKISGSILQIVNFFIAFAIYTPFFSID 428

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2215> which encodes the amino acid sequence <SEQ ID 2216>. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have no N-terminal signal sequence
60     INTEGRAL      Likelihood = -8.92      Transmembrane 347 - 363 ( 335 - 373)
      INTEGRAL      Likelihood = -7.59       Transmembrane 29 - 45 ( 27 - 50)
      INTEGRAL      Likelihood = -7.38       Transmembrane 182 - 198 ( 179 - 204)
      INTEGRAL      Likelihood = -5.68       Transmembrane 398 - 414 ( 395 - 420)
      INTEGRAL      Likelihood = -4.99       Transmembrane 293 - 309 ( 291 - 314)

```

-1106-

INTEGRAL	Likelihood = -3.61	Transmembrane	140 - 156 (134 - 160)
INTEGRAL	Likelihood = -2.60	Transmembrane	229 - 245 (229 - 245)
INTEGRAL	Likelihood = -0.75	Transmembrane	72 - 88 (72 - 88)

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 An alignment of the GAS and GBS proteins is shown below.

Identities = 366/428 (85%), Positives = 402/428 (93%), Gaps = 1/428 (0%)

Query: 1 MSKFD SQKIITPIMKPVNMRGI IALKDGLAILPLTVVGS LFLILQGLPFGKLNQAIANV 60
 M+K + Q II PIM FVNMRGI IALKDGLAILPLTVVGS LFLI GQ+PP+G+N AIA+V
 15 Sbjct: 1 MARQGMONIIPIMT FVNMRGI IALKDGLAILPLTVVGS LFLIAGQIPFGVNDIAISV 60

Query: 61 FGFTEPFPQVYSOTFAIMGLISCF AIAYAYAKNSSVEPLPAGVLSLSFFILMKSSYI 120
 FG +WTEPFMQVY GTF AIMGLISCF AI Y+YAKNS VEPLP+QVLSLS+FPIL++SSY+
 20 Sbjct: 61 FGADWTEPFPQVYHGTFAIMGLISCF AIGSYAKNSGVEPLPSGVLSLSAFFILLRSSIV 120

Query: 121 PVKGEAIDAIKSVWFGGQGIIGAIIGLVVGAITYTFIQHHIVIKMPEQVQCAIAKQFE 180
 P +GEAI DAISKVWFGGQGIIGAI+IGL VGA+YT FI+ HIVIKMPE+QVQCAIAKQFE
 25 Sbjct: 121 PABGEAIDAIKSVWFGGQGIIGAIIVIGLTVGAIVTTFIRRHIVIMPEQVQCAIAKQFE 180

Query: 181 AMIPAFVIFLLSMIVYLIANK-VITGGTFIEMIYDI IQVFLQGLTGSLYGALGIAPFFISFL 239
 AMIPAFVIF LSM+VY+IAK VT GGTFIEMIYD+IQVFLQGLTGSLYGA+GIAPFFISFL
 30 Sbjct: 181 AMIPAFVIFLLSMIVYLIANKSVGGTFIEMIYDV IQVFLQGLTGSLYGALGIAPFFISFL 240

Query: 240 WMPGVHGQSVVNGIVTALLSNLDANKSLAANRLITLNGAHIVTQQFLDSFLILSGSGI 299
 WMPGVHGQSVVNGIVTALLSNLDANK+L+AA L+LD GAHIVTQQFLDSFLILSGSGI
 35 Sbjct: 241 WMPGVHGQSVVNGIVTALLSNLDANKLMAAGSELDRGAHIVTQQFLDSFLILSGSGI 300

Query: 300 TPGLVIAMLFAAKSKQYKALGKVAAPFAIFVINEP+VGFPIVNMFMFLPILVPLVLA 359
 TPLV+AM+FAAKSKQYKALGKVAAPFA+FINNEP+VGFPIVNMFMFLPILVPLVLA
 40 Sbjct: 301 TPGLVAMIFAAKSKQYKALGKVAAPFALFINNEP+VGFPIVNMFMFLPILVPLVLA 360

Query: 360 LIVYGAIAGVFMQFPFGVTLFWSTPAIISGFMVGMQAGALVQIVILAISTAVYVFFPKIQ 419
 L VYGAIAGVFMQFP+GVTLFWSTPAIISGFMVGMQAG+VQI+IL +ST VYVFFPKIQ
 45 Sbjct: 361 LTVYGAIAGVFMQFPAGVTLFWSTPAIISGFMVGMQAGVQIILIMSLTVYVFFPKIQ 420

Query: 420 DNITYKNE 427
 DN+ Y+NE
 Sbjct: 421 DNMAYQNE 428

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1000

A DNA sequence (GBSx1060) was identified in *S. agalactiae* <SEQ ID 3067> which encodes the amino acid sequence <SEQ ID 3068>. This protein is predicted to be formate acetyltransferase 2 (pflB). Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5049(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

60 >GP:AACT73910 GB:AB000184 putative formate acetyltransferase

-1107-

[Escherichia coli K12]
 Identities = 414/805 (51%), Positives = 555/805 (68%), Gaps = 14/805 (1%)

Query: 25 LTERWVSYRDKVLD-KKPFDAERAILVTRAYQHQKFNVLKRAVMLQNLEKMTYIID 83
 L++R+ ++++ ++ KP + ERA TS YQ+H +KP ++RA L + L TI+I
 Sbct: 9 LSRILKAHKNAGVHTVKPPVCTERAQHYTMYQQHLDKPIPVKRALALAHLANRTINIK 68

Query: 84 DETMIVNGQSSDKDAIPPEYTLLEPVNVLDELPEKRDGDVPTITETKQIRINIAFPW 143
 ++I+QWAS + APIFPKYTA ++ E+D R G F ++EH K + + P+W
 Sbct: 69 HDELIIGNQASEVRAAPIPKYTVSWIEKIDDLADPGAGFVSEKMRVLHEVCFWWR 128

Query: 144 NNNLRARAGVMLPEEVQVYMFETGFGHGNMNSGDAHLAVNYQKLEBGLIGPEKKARKA 203
 ++ R M +E + + TG EG M SGDHLAVN+ LLE+GL G ++ +
 Sbct: 129 GQTVQDRCYGMFTDEQKSLATGIIKAGNMTSGDAHLAVNPFLLEKGLDGLREVAER 188

Query: 204 KADLDLTKEPSIDKYHFPYDSILITIKAVKTYARRFAILAKKQAKTANAK-RRQIMLDIAS 262
 ++ ++LT E + F + I I + AV + ERFA LA++ A T + RR ELL +A
 Sbct: 189 RSRINLTVLEDLHGEQFLKALDIVLAVSEHIERFAALAREMAATETRESRDELLAMAS 248

Query: 263 ICERVPVYPAETFAENVSQVWFQITQLQIESNGHSLSYGRFDQYMYVYKSLPEKRW 322
 C+ + + P +TF +A+Q +FIQ ILQIESNGHS+S+GR DQY+YYP +D+E + +
 Sbct: 249 NCDLIAHQPCQTFWQALQLCYFIQLILQIESNGHSVSFRDQYIYYPYTRRDVNLQITD 308

Query: 323 -DSVERLTNMLIKTITINKVRSQHTPSSAGSPLYQNVTIGGQTR---HKEDAVNPLSF 378
 + +E L + W+K + +NK+RS +H+ +SAGSPLYQNVTIGGQ DAVNPLS+
 Sbct: 309 REHATEMLHSCMLKLEVNKIRSSSHSKASAGSPLYQNVTIGGQNLVDQPMQDAVNPLSY 368

Query: 379 LVLSVAQTHLPQNLITVRYHANLKSFMNEAIEVMKLGFGMPAFNNDRIIIPSPIKRW 438
 +L+S + QNML+VRYHA + P++ ++V++ GFGMPAFNNDRI+IP PTK G+
 Sbct: 369 AILESQGLRSLSTQNLISVRYHAGSNDFLDACVQVIRCGFMPAFNNDRIEVIPEIKLGI 428

Query: 439 SEEDAYDYAIGCVETAVGKNGVYRCTGMSYINFVKVLLITMNDGIDPASGRKRP--- 494
 +DAYDY+AIGC+ETAV GKGNGVYRCTGMS+INF +V+L + G D SGK F P
 Sbct: 429 EPQDAYDYAIGCIETAVGKNGVYRCTGMS+INFARVMLAALGGHDATSGKVFLOPEKA 488

Query: 495 -SYGHFTQWTSYKELKEAMDKTLRYLTRMSVIVENAIDISLEREVPDILSALITDCIGR 553
 S G+F ++ E+ +AMD +RY TR S+ +E +D LE V DILCSAL DDCI R
 Sbct: 489 LSAQNFN---NPDEVMDADTQIRYTYRKSIEIEVVDTMLENVHIDILCSALVDCTIER 545

Query: 554 GKHLKEGGAVDYISGLQVGIANLSLSLA+KKLVFEKKRLTLEWVQALQSDYAGPRGE 613
 K +K+GGA YD++SGLQVGIANL+SLAA+KKLVFE+ + + AL D+ G E
 Sbct: 546 AKSINQGAKYDWSGLQVGIANLSLA+KKLVFEQGAIGQQQLAALADDFGLTGE 605

Query: 614 BIRQMLINEAFKYNDDVDVSLVRECYDVVREIAKYPNTRYGRGP+IGGIRYSGTSSIS 673
 ++RQ LIN APKYGNDD D+L+ Y Y++E+ +Y N RYGRGP+GG Y+GTSSIS
 Sbct: 606 QLRQLINGAFKYNDODTVDTLLARAYCTYIDELKQYHNPRYGRGFGVGGNYTAGTSSIS 665

Query: 674 ANVQCGROTLATPQGRHAGTPIAAGCSPSHNMDDKGPISVLKSVSKLPDEIVGGVLNLQ 733
 ANV G T+ATPQGR A TPIAAG SP+ D GPT+V+ SV KLPT I+GGVLNLQ
 Sbct: 666 ANVPFGAQTATPQGRKAHTPIAAGASPASGTDHGPATVIGSVSKLPTAAILGGVLNLQ 725

Query: 734 K+NPCTLAKEDKQKLIALLRTFFNHLGSHYIQYNNVVSRETLIDAKHPEKHRDLIVRVA 793
 K+NP TL E DKQKL+ LRTEFF G+H+IQYV+VSRETL+DA+KHP++RDL+VRVA
 Sbct: 726 KLNPATLENESDKQKLMILLRTFFEVHKGWHIYQYNNVVSRETLIDAKHKGQYQVLDIVRVA 785

Query: 794 GYSAPFNVLKATQDDIIARTHEAL 818
 GYSAPF LS QDDIIARTHE L
 Sbct: 786 GYSAPFTALSPDAQDDIIARTHEML 810

60 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3069> which encodes the amino acid sequence <SEQ ID 3070>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

65 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4763 (Affirmative) < succ>

-1108-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5	Identities = 694/803 (86%), Positives = 747/803 (92%)
	Query: 16 QNSQKHGPHILTERMYSYRDKVLKKPFIIDAEKAILVTERYQKHQKPKNVLRKRAYMLQNIL 75
	+ +PG+LT+RM YR+ VLDKKP+IDAERAIL TEAYQKHQ KP LKRAYMLQ IL
10	Sbjct: 3 ETKSPYFPHGLIDRTHCYREAVLDKKPYIDAERAILATEAYQKHQKPNLKRAYMLQTL 62
	Query: 76 ENMTIYIDBTMIYGNQASSDKDAIFPEYTLBFWVKLDFEKRKGQVFIYTEETKQI 135
	E MTIYI+DE++I GNQASS+KDAIFPEYTLBFW+NELDLFEKRKGQVFIYTEETK+Q+
	Sbjct: 63 ENMTIYIEDESLIAGNQASSNKDAIFPEYTLBFWVKNELDLFEKRKGQVFIYTEETKQQL 122
15	Query: 136 RNIAPFWERNRLRARAGVMLPEEVQVYMETGFFGMGEGKNSGDHIAVNYQKLEEGIG 195
	R+IAPFWERNRLRAR GV+LPEEVQVYMETGFFGMGEGKNSGDHIAVNYQKLE GL G
	Sbjct: 123 RDIAPFWERNRLRARCGVLLPEEVQVYMETGFFGMGEGKNSGDHIAVNYQKLEHGLKG 182
20	Query: 196 FEKARKAKADLLDTKYPESIDKHFYDS ILITIEAVRTYAEFALIAKQAKTANAKRRQ 255
	FE++AR AKA LDET PE+IDKHFYDS+ I I+AVRTYA+R+A LA++ AKTA +R+
	Sbjct: 183 FEERARAAGALLDITIPENIDKHFYDSVFVIDAVRTYAKRYAKLAEAKTAKPERQA 242
	Query: 256 ELLDIASICERVFYPAETFAEAVQSVWFICQLQIESNGHSLSYGRDQYMYPTVVKDL 315
	ELLDIA ICA+VPY PA+TFAEAVQSVWFICQLQIESNGHSLSYGRDQYMYPTVVK+DL
25	Sbjct: 243 ELLDIARICDKVFPYPAETFAEAVQSVWFICQLQIESNGHSLSYGRDQYMYPTVVKDL 302
	Query: 316 EAGRETEDSIVERLITMLWIKTITINKVRSQAHTFSSAGSPLYQNVITIGGCTRHDEAVNP 375
	EAGRETED+IVERLITMLWIKTITINKVRSQAHTFSSAGSPLYQNVITIGGCTR K+DAVNP
	Sbjct: 303 EAGRETEDTIVERLITMLWIKTITINKVRSQAHTFSSAGSPLYQNVITIGGCTRDKDAVNP 362
30	Query: 376 LSPVLVLSVAQTHLPQMLTVRYHANLOKSPMNEAIKMLGFGMPAPNDEIIPSPFK 435
	L+S+LVL+VQAQT LPQMLTVRYH LD +PME IEVMLGFGMPA NDEIIPSPFK
	Sbjct: 363 LSYLVLSVAQTKLPQMLTVRYHGLDWTPEMEIEVMLGFGMPAPNDEIIPSPFK 422
35	Query: 436 KGVSEEDAYDYSAGCVETAVPGKMGYRCTGMSYINFPKILLITMDGIDPASGKRFAP 495
	KGVSEEDAYDYSAGCVETAVPGKMGYRCTGMSYINFPK+LLITMDGIDPASGKRFAP
	Sbjct: 423 KGVSEEDAYDYSAGCVETAVPGKMGYRCTGMSYINFPKILLITMDGIDPASGKRFAP 482
40	Query: 496 YGHPTQMTSYKELKEAWDKTLRYLTRMSVIVENADISLERVDPDILCSALTDGICGRK 555
	+GHP HTSY+ELK AND TLR +TRMSVIVENAD+ LERERVDPDILCSALTDGICGRK
	Sbjct: 483 HGHFKDMTSYELKLAAMDATLKEITRMSVIVENADILGLERERVDPDILCSALTDGICGRK 542
	Query: 556 ELKGGAVYDYISGLQVGIANLSDSLAALKCLVFEERLITLVEVQALQSDVAPGRBEI 615
	LKGGAVYDYISGLQVGIANLSDSLAALKCLVFEER LIT E+W+AL+SD+AG RGE+I
45	Sbjct: 543 TLKGGAVYDYISGLQVGIANLSDSLAALKCLVFEERLITLVEVQALQSDVAPGRBEI 602
	Query: 616 RQMLINAPKYGNDDYADSLVRECYDVVVEIAKYVNTYRGRFGIGIRYSGTSSISAN 675
	RQMLIN+APKYGNDDYADSLV E YD Y++EIAKYVNTYRGRFGIGIRYSGTSSISAN
	Sbjct: 603 RQMLINDAPKYGNDDYADSLVREYDYTYDEIAKYVNTYRGRFGIGIRYSGTSSISAN 662
50	Query: 676 VQGGKGTLATPDGRHAGTPLABGCSFPHNMOKKGTSPVLKSVSKLPTEIVGVVLINQKV 735
	VQGG+GTLATPDGRHAGTPLABGCSF H+MOKKGTSPVLKSV+KLPTEIVGVVLINQKV
	Sbjct: 663 VQGGKGTLATPDGRHAGTPLABGCSFPHNMOKKGTSPVLKSVAKLPTEIVGVVLINQKV 722
55	Query: 736 NPQTLAKREDKIKLALRTFFPNRLGHYHIQYNNVSRRLTDNQKHPEKRRHLIRVAGY 795
	NPQTLAKREDK KL+ALLRTFFPNRLGHYHIQYNNVSRRLTDNQKHPEKRRHLIRVAGY
	Sbjct: 723 NPQTLAKREDKIKLALRTFFPNRLGHYHIQYNNVSRRLTDNQKHPEKRRHLIRVAGY 782
	Query: 796 SAFFNVLSKATQDDIIARTEHAL 818
	SAFFNVLSKATQDDII RTEH L
60	Sbjct: 783 SAFFNVLSKATQDDIIARTEHTL 805

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1001

A DNA sequence (GBSx1061) was identified in *S.agalactiae* <SEQ ID 3071> which encodes the amino acid sequence <SEQ ID 3072>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1024 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.
>GP:CAA05516 GB:AJ002527 OxFX [Clostridium beijerinckii]
Identities = 90/214 (42%), Positives = 131/214 (61%), Gaps = 1/214 (0%)

Query: 1 MEFLDTLNLEAIKQWHILPLAGVTSNPITAKKBGDIHFQRIQIRDVREIGREASLIHQ 60
M+ ++D +N+E IK I + GVTSNP+I K G + I+ +RE IG + LHVQ
Sbjct: 1 MKLIDVDVNIKIKIDVRSIFQIDGVTSNPSILHKYGKQPYBILII-IRFFIGENSLHVQ 59

Query: 61 VVAKDYQGILDDAAKIRQETDDDIYIKVPVTPDGLAAIKTLKAEQYNTATATAYTSMQGL 120
V+++ +G+L +A KI +E + Y+K+PVT DGL AIK L+ E N+TATAYT MQ
Sbjct: 60 VISSESGMLKEAHKIIKELGKNTYVKIPVTRDGLKAIKILRKEBINVTATAYTQMAY 119

Query: 121 LAISAGADYLAIFYNNRMENLIDATQVIKELAQAIERTGSSSKILAAFPKASQVTKALS 180
LA AGA Y APY NR++NL + QV K++ E+ +++LAASFKN+ QV +
Sbjct: 120 LAGKAGAQYAAPVYNRIDNLGANGVQAKDIDHIFENKPFTEVLAASFKNSQQVLELCK 179

Query: 181 QGAQSITAGPDIFESVFAMPSTAKAVNDFADDMK 214
G + T ED+ E+ + AV +F D++
Sbjct: 180 YGIGATISPDVIEGLIKNDICVDVAVENPKDFE 213

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3073> which encodes the amino acid sequence <SEQ ID 3074>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1090 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 158/222 (71%), Positives = 194/222 (87%)

Query: 1 MEFLDTLNLEAIKQWHILPLAGVTSNPITAKKBGDIHFQRIQIRDVREIGREASLIHQ 60
ME++LDTL+LEAIKQWHILPLAGVTSNP+IAKKBG+I FF+RIR+VR IIG +AS+HVQ
Sbjct: 1 MEYMLDTLDEAIKQWHILPLAGVTSNPSTAKKBGDIHFPERIREVRAIGDKASLIHQ 60

Query: 61 VVAKDYQGILDDAAKIRQETDDDIYIKVPVTPDGLAAIKTLKAEQYNTATATAYTSMQGL 120
V+A+DY+GLD DAA+R++ D +Y+KVPVT +GLAAIKTLKAEQY+ITATAYT+ QGL
Sbjct: 61 VIAQYVYGLIKLDAEIRQCGDSVYKVPVTPDGLAAIKTLKAEQYHITATAYTTFQGL 120

Query: 121 LAISAGADYLAIFYNNRMENLIDATQVIKELAQAIERTGSSSKILAAFPKASQVTKALS 180
LAI AGADYLAIFYNNRMENL+ID VI++LA+AI R ++SKILAAFPK+ QV K+ +
Sbjct: 121 LAIEAGADYLAIFYNNRMENLIDPRAVIEQLAATNRKRNANSKILAAFPKVAQNKPSA 180

Query: 181 QGAQSITAGPDIFESVFAMPSTAKAVNDFADDMKASQHSKI 222
GAQ+ITAGPD+FE+ FAMPSTI KAV+DFG DW+A H + I
Sbjct: 181 LGAQAITAGPDVFEAGFAMPSTIQKAVDFGDKWRAIHRKSI 222

```

-1110-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1002

A DNA sequence (GBSx1062) was identified in *S.agalactiae* <SEQ ID 3075> which encodes the amino acid sequence <SEQ ID 3076>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3086 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9545> which encodes amino acid sequence <SEQ ID 9546> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GF:AAA22477 GB:M65289 glycerol dehydrogenase [Bacillus
  stearothermophilus]
Identities = 199/362 (54%), Positives = 271/362 (73%), Gaps = 2/362 (0%)

Query: 4  KVFASPSRYIQGKDALFQSIIEHIKSLGQTLPLICDDVVYVNIQVGRFLSYLOD-DLLPHRV 62
+VF SP++Y+QGG+ + + +++ +G +++ D++V+ I G ++ Lt ++ V
Sbjct: 5  RVFISPAKYVQGGKVNITKLANVLEIGINKTVVLADEIVMKIAGHTIVNLELKGNIIEAEV 64

Query: 63  SFNGEASDNEINRVVAVAKKNSDLIIQLGGKTTDSAKAIADKYNRPVIAPTVASTDA 122
F+GEAS NE+ R+ +A++ + ++IG+GGKT+D+AKA+AD++ +VI PT ASTDA
Sbjct: 65  VFSGEASRNEVERIANLARKAAEAVIGVGSGKTLDTAKAVALDELDAIVIVPTASTDA 124

Query: 123 PTSALSVIYTDGEAFKYIFYSKNPDLVLDTCVIAQAPKRLASGIADLATVVEARV 182
PTGALSVIY+D+G FE Y FY KNPLDLVLD+IA AP RLLASGIAD LATVVEAR+V
Sbjct: 125 PTSALSVIYSDDGVFESYRFYKKNPDLVLDTKIIANAPRLLASGIADALATVVEARSV 184

Query: 183 LQNGIAGVAGGRQTLGAVIAQACERTLNDLQALAACDAKVVTKALENVIENTLLSG 242
++ G MAGG T+A AIA+ CE+TLF A + AKVVT ALE V+ENTLLSG
Sbjct: 185 IKSGGKTWAGGIPTIAAEATAEKCEQTLFKYKLAYESVKAKVVTFALEAVVENTLLSG 244

Query: 243 LGFESAGLAAAHAIHNGFTALSGDIHHLTHGEKVAYGTLTQLFLENRPKEIDRYINLYQ 302
LGFES GLAAAHAIHNGFTAL G+IHHLTHGEKVA+GTL QL LE ++EI+RYI LY
Sbjct: 245 LGFESGGLAAAHAIHNGFTALSGEIHHLTHGEKVAFGTLVQLALEESQOEIRYIYLYL 304

Query: 303 AIGMPTTLAELHGDATYESLLKVGQQAATGETHIEMFPKISAEVDVAALLTVDRYSN 362
++ +P TL ++ L DA+ E++LKV + AT EGETH F ++A+DVA A+ D+Y
Sbjct: 305 SLDLFTVLEDIKLKASREDILKVAKAATAGSETIHN-AFNVATDQVADAFADQYAKA 363

Query: 363 HQ 364
++
Sbjct: 364 YK 365
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3077> which encodes the amino acid sequence <SEQ ID 3078>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -4.62 Transmembrane 101 - 117 ( 98 - 119)

----- Final Results -----
bacterial membrane --- Certainty=0.2848 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```


The protein has homology with the following sequences in the databases:

>GP:AAA22477 GB:M65289 glycerol dehydrogenase [Bacillus
stearothermophilus]
1 Identities = 202/357 (56%), Positives = 261/357 (72%), Gaps = 1/357 (0%)
Query: 2 KVFASPSRYIQGKALFTNVKILQGLQDSPIILLCDVVYIGVGERFESYLIDNGMTPVHV 61
++VP SP++Y+QGNK + L+ +G+ +++ D++V+ I G + L + + V
10 Sbjet: 5 RVFISPAKYVQGNKIVITKIANYLGGIKGNKTVVIADETVMKIAHGHTVNEIKKKNIAAREV 64
Query: 62 AFNGEASDNEISRVVAIAKENGNDVILGLGGKTTDSAKAIADLAVPVIITPTASTDA 121
F+GEAS NE+ R+ IA++ ++IG+GGKTD+D+AKA+AD L + +I PT ASTDA
12 Sbjet: 65 VFGGEASRNEVERIANIARKEAAIIVIGVGGKTLDTAKVADELDAYIVIVPTASTDA 124
Query: 122 PISALSVIYTDGGAPEKFIYFSKNPDLVLVDTVQVQAPKRLLASGIADGLATWVEARV 181
PISALSVIY+D+G FE Y FY KNPDVLVNDT++I AP RLLASGIAD LATWVEAR+V
15 Sbjet: 125 PISALSVIYSDGVPFESYRFYKKNPDLVLVDTKIANAPRLLASGIADGLATWVEARSV 184
Query: 182 MQKNGDTMAGGNQTLAGVAIAKACEQTLFADGLKAMASCDROVVTPALENVIEANTLLSG 241
++ G TNAGG T+A AIA+ CEQTLF G A S +VVTPALE V+EANTLLSG
20 Sbjet: 185 IKSOGKTMAGGIPITAAABAIEKCEQTLFKYKGLAYESVKAKVVTPALEAVVEANTLLSG 244
Query: 242 LGFBSAGLAAAHAIHNGFTALGTGAIHLTHGEKVAYGTLTQLFLENRSREEDIRYIDFYQ 301
LGFBS GLAAAHAIHNGFTAL G IHLTHGEKVA+GTL QL LE S++EI+RYI+ Y
25 Sbjet: 245 LGFBSAGLAAAHAIHNGFTALGEGTHHLTHGEKVAFGTLVOLALEHSQBIERYITELYL 304
Query: 302 AIGMPTTLKEMHLDATQEDFLKIGRQATMAGETTHQMPPVISPEDVAALVAVDAY 358
++ +P TL+++ L A++ED LK+ AT GETH F ++ +DVA A+ A D Y
30 Sbjet: 305 SIDLPVTELEDKLDASREDILKVAKATAGETHFN-AFNVITADDVADAIADADQY 360

An alignment of the GAS and GBS proteins is shown below.

Identities = 287/361 (79%), Positives = 325/361 (89%), Gaps = 1/361 (0%)
Query: 3 MKVFASPSRYIQGKDALFQSIHFKSLGQTPILLCDVVYINIGVERFESYLQDD-LLPHR 61
MKVFASPSRYIQGK+ALF +++ +K LG +P++L+CDVVY IUGERF SYL D+ + F
35 Sbjet: 1 MKVFASPSRYIQGNALFTNVKILQGLQDSPIILLCDVVYIGVGERFESYLIDNGMTPVH 60
Query: 62 VFPNGEASDNEINRVVAIAKEKNDLILGLGGKTTDSAKAIADKVLNPVVIPTASTD 121
V+PNGEASDNEI+RVVA+AKE +D+IIGLGGKTTDSAKAIAD + +PV+IAPT+ASTD
40 Sbjet: 61 VAFNGEASDNEISRVVAIAKENGNDVILGLGGKTTDSAKAIADLAVPVIITPTASTD 120
Query: 122 APTSALSVIYTDGGAPEKFIYFSKNPDLVLVDTVQVQAPKRLLASGIADGLATWVEARA 181
APTSALSVIYTDGGAPEKFIYFSKNPDLVLVDTVQV QAPKRLLASGIADGLATWVEARA
45 Sbjet: 121 APTSALSVIYTDGGAPEKFIYFSKNPDLVLVDTVQVQAPKRLLASGIADGLATWVEARA 180
Query: 182 VLQKNGIANGAGGROTLAGVAIAQACERTLFPNDLSQALAAACDQVVTALENVIEANTLLS 241
V+KNG MAGG QTLAGVAIA+ACE+TLF D L+A+A+CD +VVT ALENVIEANTLLS
50 Sbjet: 181 VMQKNGDTMAGGNQTLAGVAIAKACEQTLFADGLKAMASCDROVVTALENVIEANTLLS 240
Query: 242 GLGFBASGLAAAHAIHNGFTALGTGAIHLTHGEKVAYGTLTQLFLENRPKEEDIRYINLY 301
GLGFBASGLAAAHAIHNGFTAL+G IHLTHGEKVAYGTLTQLFLENR +EEDIRYI+ Y
55 Sbjet: 241 GLGFBASGLAAAHAIHNGFTALTGAHILTHGEKVAYGTLTQLFLENRSREEDIRYIDFY 300
Query: 302 QAIQMPTTLAEHLGADATYEELKVGQQTATIGETTHMPFKISAKIVAAALVVDKRYVSN 362
QAIQMPTTL E+HL AT E+ LK+G+QAT+ GETTH+MPF IS EDVAAAL+ VD YV++
Sbjet: 301 QAIQMPTTLKEMHLDATQEDFLKIGRQATMAGETTHQMPPVISPEDVAALVAVDAYVTS 361

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1112-

Example 1003

A DNA sequence (GBSx1063) was identified in *S.agalactiae* <SEQ ID 3079> which encodes the amino acid sequence <SEQ ID 3080>. Analysis of this protein sequence reveals the following:

```

Possible site: 28
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.75    Transmembrane    262 - 278 ( 262 - 279)

----- Final Results -----
      bacterial membrane --- Certainty=0.1298 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAAB8310 GB:AB028865 O-acetylserine lyase [Streptococcus suis]
Identities = 239/304 (78%), Positives = 273/304 (89%)

Query: 4 IYNSITDLIGNTPIQLHIVPFGAAEVYVKLESFNPQSSVKDRIALAMIEDAQKGLK 63
      IY +IT LG TP +I+L+-IVPFGAAEVYVKLS+FNPGSSVKDRIALAMIEDAQ+ G +K
Sbjct: 3 IQYNITQLWGKTPVVKLNINIVPFGAAEVYVKLEAFNPGSSVKDRIALAMIEDAKAGTTK 62

Query: 64 AGDTIVPSTSGNTGIGLAWVGKANGYVNVIIIMPETMSIERKKIIQAYGAQLVLTGPGSEG 123
      GDTIVPSTSGNTGIGLAWVG KANGYVNVIIIMPETMS+ERRKIIQAYGA+LVLTGPGSEG
Sbjct: 63 PGDTIVPSTSGNTGIGLAWGAKGANGYVNVIIIMPETMSVERKKIIQAYGAELVLTGPGSEG 122

Query: 124 KGALAKAKEISAEQNWLFLQFNQANPEIHEKITYGRBIETFGKEKGLDAPFAGVGTGT 183
      KGALAKAKEI+ E+N W+P OF N +NP++HE TTG+BI+E FG GLDAF++GVGTGTGT
Sbjct: 123 KGALAKAKEIAEEKNGWVFPQFANPNPKVHEDTTGQRILEDFTGTGLDAFVSGVGTGTGT 182

Query: 184 ITGVSRLKKNVNDVAIYVAEDESAILSGEOPGPHKIQGISAGFIPETLATDSYDIIR 243
      ++GVS LK NPD+AIYVAEDESAA+LSGE PGPHKIQGISAGFIP+TL T +YD IIR
Sbjct: 183 VSGVSHVLTANPDIAIYVAEDESAILSGEAPGPHKIQGISAGFIPDTLTSYDGIIR 242

Query: 244 VTSDDAETGRIIGGLEFLAGISASAAIYAAIEVAQLGKGVKVLALLPONGERYLST 303
      V SDDA+ TGR IGG EGF. GIS+ AAI+AAIEVAK+LG GKKVLA+LPONGERYLST+
Sbjct: 243 VKSDDALATGRAIGGEGFLVGISSGAATHAAIEVAELGTGKKVLAALLPONGERYLSTA 302

Query: 304 LYDF 307
      LY+ F
Sbjct: 303 LYEF 306

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3081> which encodes the amino acid sequence <SEQ ID 3082>. Analysis of this protein sequence reveals the following:

```

Possible site: 58
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.48    Transmembrane    262 - 278 ( 262 - 278)

----- Final Results -----
      bacterial membrane --- Certainty=0.1192 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:BAAB8310 GB:AB028865 O-acetylserine lyase [Streptococcus suis]
Identities = 235/303 (77%), Positives = 261/303 (85%)

Query: 4 IYXTITELVGQTPFIKLNRLPNKADVYVVKLEAFNPGSSVKDRIALSMIEAAEASGLIS 63
      IY+ IT+LVG+TP+IKLN++P AA+VYVKLEAFNPGSSVKDRIAL+MIE AE G I I
Sbjct: 3 IQYNITQLWGKTPVVKLNINIVPFGAAEVYVKLEAFNPGSSVKDRIALAMIEDAQKAGTTK 62

Query: 64 PGDVIIEPTSGNTGIGLAWVGKANGYVNVIIIMPETMSIERKKIIQAYGAQLVLTGPGSEG 123
      PGD I+EPTSGNTGIGLAWVGKANGY VVIMPETMS+ERR+IIQAYGAELVLTGPG+EGM

```

-1113-

Sbjct: 63 PGEIVPTSGNIGIGLAWGAAGKYNVIIVMPETMSVERRKI:QAYGAELVLTGSEGM 122

Query: 124 KGAIKAETLAIELGAWMPQFNNPANPSIHEKTTAQEILEAPKEISLDAFVSGVGTGT 183
KGAIKA+ +A R W+P OF NP+NP +HE TT QREIL F LDAFVSGVGTGT

5 Sbjct: 123 KGAIKAKEIAKEKNGWVPQFANPNPKVHED:TGQRILEDFTGTGLDAFVSGVGTGT 182

Query: 184 LSGVSHVLKKNPVTIYAVEAESAVLSGQEPGPHKIQGISAGFIPNLTLDKAYDQIR 243
+SGVSHVLK ANP+ IYAVEA+ESAVLSG+ PGPHKIQGISAGFIP+TLTD AYD IIR

10 Sbjct: 183 VSGVSHVLKIANPDIAIYAVEAESAVLSGQEPGPHKIQGISAGFIPNLTLDKAYDQIR 242

Query: 244 VKSKDALETARLTGAKEGFVIGISSGAALYAAIEVAKQLGKGHVLTILPDNGERYLST 303
VKS DAL T R G KEGFVIGISSGA++AAIEVAK+LG GK VL ILPDNGERYLST

Sbjct: 243 VKSDDALATGKRAIGKEGFVIGISSGAALYAAIEVAKQLGKGHVLTILPDNGERYLST 302

15 Query: 304 LYD 306
LY+

Sbjct: 303 LYE 305

An alignment of the GAS and GBS proteins is shown below.

20 Identities = 222/306 (72%), Positives = 263/306 (85%)

Query: 1 MSKIYNSITDLIGNFTIQLHHIVPEGAEEVYKLESFNPSSVKRIALNMIEDAEKG 60
M+KIY +IT+L+G TPIL+L+ ++P AA+VYKLE+FNPOSSVKRIAL+MIE AE +G

25 Sbjct: 1 MTKIYKTTITELGQTFI:KINRLIPNEADVYKLEAFNPSSVKRIALSMIEAAEAG 60

Query: 61 ILKAGDTIVETSGFTIGLAWVSGAKGYNVIIVMPETMSIERRKLIQAYGAQLVLTGGS 120
++ CD.I+ETSGFTIGLAWVSG AKGY VIIVMPETMS+ERR+LIQAYGA+LVLTG+

30 Sbjct: 61 LISPGDVIIETSGFTIGLAWVGAAGKYNVIIVMPETMSLERRQIIQAYGAQLVLTGGA 120

Query: 121 EGNKGAIKAKEIASQANWMLPLCFNNQANPEIHEKTTIGREIITETPGEKGLDAFIAGVGT 180
EGNKGAIKA+ ++ E AW+P+QENN ANP IHEKIT +RI+E F E LDAF+LVGT

35 Sbjct: 121 EGNKGAIKAETLAIELGAWMPQFNNPANPSIHEKTTAQEILEAPKEISLDAFVSGVGT 180

Query: 181 GGTITOVSRALKKVPDVAIYAVEAESAVLSGQEPGPHKIQGISAGFIPETLATDSDYH 240
GTT+GVS LKK NP+ IYAVEA+ESA+LGS+PGPHKIQGISAGFIP TL T +YD

40 Sbjct: 181 GGTLSGVSHVLKKNPVTIYAVEAESAVLSGQEPGPHKIQGISAGFIPNLTLDKAYDQ 240

Query: 241 IIRVTSDDALETGRITIGLEGFVIGISSGAALYAAIEVAKQLGKGHVLTILPDNGERYL 300
IIRV S DA+ET R+ G RGFV GIS+ AA+YAAIEVAKQLGKGHV L+LPDNGERYL

45 Sbjct: 241 IIRVSKDALETARLTGAKEGFVIGISSGAALYAAIEVAKQLGKGHVLTILPDNGERYL 300

Query: 301 STSLYD 306
ST LYD

Sbjct: 301 STELYD 306

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1004

50 A DNA sequence (GBSx1064) was identified in *S.agalactiae* <SEQ ID 3083> which encodes the amino acid sequence <SEQ ID 3084>. Analysis of this protein sequence reveals the following:

Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.3666(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

60 >GP:BAR07349 GB:AP001519 unknown conserved protein [Bacillus halodurans]
Identities = 96/204 (47%), Positives = 127/204 (62%)

-1114-

Query: 2 NYTKIKSDGIVEEIKKSRFICHLKRVSESEGRNYITQIKKHAHYKANHSCSAMVIGEGK 61
 +Y T+K GI E I+KSRFI HL R SREE +I QIKK H+ A H+CSA +IGE
 Sbjct: 4 SYTTVKESGIHEISIQKSRFIAHLSEATSEBKAQTFEQIKKEHWNATNCAYLIGEND 63

Query: 62 DIKRSSDDGEPSTAGIPMLTVLEKQGLTNVAVVIRYFGGKLGAGGLIRAYSGSVANT 121
 ++++DDGEPSTAG+PML VL+K+ L + VAVVIRYFGG+KLGAGGLIRAY +V++
 Sbjct: 64 QVQKANDGEPSTAGVPMLEVLKKRRLLKDTVAVVIRYFGGKLGAGGLIRAYSGSVANDG 123

Query: 122 IKRIGVVEVKEQIGIRIQLTYPOVQTFDNFLKEHHLQRFETEFLEAVTCKIYVDPKEFEH 181
 + IGUVK K I + Y +N L+ H E +LE V + YV E E
 Sbjct: 124 LNAIGVVERKRMQVHTSIDYHMLGKVNEELQSHYLLKEISYLENDVDQTVYLEAVES 183

Query: 182 TITNLTEFYQGKALLTEBGSQIVE 205
 +T G+A T + +E
 Sbjct: 184 YCEWMNLINQQAAPTHGAIBYLE 207

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3085> which encodes the amino acid sequence <SEQ ID 3086>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.43 Transmembrane 86 - 102 (86 - 102)

----- Final Results -----
 bacterial membrane --- Certainty=0.1171 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9153> which encodes the amino acid sequence <SEQ ID 9154>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.43 Transmembrane 81 - 97 (81 - 97)

----- Final Results -----
 bacterial membrane --- Certainty=0.1171 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 122/206 (59%), Positives = 153/206 (74%)

Query: 2 NYTKIKSDGIVEEIKKSRFICHLKRVSESEGRNYITQIKKHAHYKANHSCSAMVIGEGK 61
 ++ETIK+ G EE IKKSRFICH+KRV +EE+G+N+ IKK HYKANHSC AM+IG
 Sbjct: 8 HFKTIKAGFFESIKKGRFICHLKRVSTEDGKNFVNAIKKHYKANHSCFAMLIIGNR 67

Query: 62 DIKRSSDDGEPSTAGIPMLTVLEKQGLTNVAVVIRYFGGKLGAGGLIRAYSGSVANT 121
 IKRSSDDGEPSTAGIP+L+VLEKQ LTNV VVIRYFGGKLG GGLIRAYS A
 Sbjct: 68 DIKRSSDDGEPSTAGIP+L+VLEKQGLTNVAVVIRYFGGKLGGLIRAYSNMATA 127

Query: 122 IKRIGVVEVKEQIGIRIQLTYPOVQTFDNFLKEHHLQRFETEFLEAVTCKIYVDPKEFEH 181
 IK G++EVK+QIG+ I L+YFCQY + N L + L E ET+F + + +Y D + B+
 Sbjct: 128 IKRPGIIEVKQIGLGLITLSPYQYOLYSNLLDQALTEETETKPSDTIKTLTYCDTERVEN 187

Query: 182 TITNLTEFYQGKALLTEBGSQIVEIP 207
 I LT +Y G+ + GS++E P
 Sbjct: 188 LIDTLNYYHGQISCEKIGSKVIEFP 213

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1005

A DNA sequence (GBSx1065) was identified in *S.agalactiae* <SEQ ID 3087> which encodes the amino acid sequence <SEQ ID 3088>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1421 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA04940 GB:U56901 involved in transformation [Bacillus subtilis]
 Identities = 160/405 (39%), Positives = 228/405 (56%), Gaps = 20/405 (4%)
 Query: 35 YICTRCSSSVAKNQQL---PTGNYCYCRGCIQVPRVTSNENLYYFPKQTFKTSNLSK--W 88
 Y C R C + + Y C R C + GRV+ LY + + + S + K W
 Sbjct: 58 YRCRCGQTDQRYPSFYHSSGKNKLYCRSCVMGRVSEEVPLYSWKEENESNMKSIKLW 117
 Query: 89 KGEITPYQNEVSEELLKGTSSKENLLVHAVTGAGKTEMIYHSAVKVIDTGGSVCIASPRI 148
 G+L+ Q + + L+ + IS KB LL+ AV GAGKTE++ + ++ G VCI+PR
 Sbjct: 118 DGLSSGQQAANVLEIAISKKEELLINAVCGAGKTEMLPGIESALNQGLVCIATPKT 177
 Query: 149 DVCLLEYKRLSDFRCA-ITLMHGSPSYQR-SPLTIATTHOLLKFYHAFDILLIVEVDA 206
 DV LEL RL P+ A I+ ++G S R SPL I+TTHOLL++ A D++I+DEVDA
 Sbjct: 178 DVVLEIAPRIKAAPOQADTSALYGSISDCKGRISPLMISTTHQLRYKDAIDVMIIDEVDA 237
 Query: 207 PPVVDNPILYGVKQALENGTSIFLTATSTELERKARKELKKLHARRFHANPLVIP 266
 PPY + L V++A K+N T ++L+AT EL+RK +L + + R H PL P
 Sbjct: 238 PPYSADQTLQFAVQYARKKNSTLYVLGATPPEKLLKRAKALNQHSVRIPARHNRKPLPSP 297
 Query: 267 ENVVSIGIKSQSLQQLPKPKLYQLINKQRTYRPLLLFPFHISBQVFTILRQAFPMK 326
 VW +K L K+PP + + I + P+ LFP+S IL +A K
 Sbjct: 298 RFWCGNMKKKLNRNKIPPAVKRWIEFHVKEGRFVFLFVPSV-----ILEAAACPK 350
 Query: 327 -----IGFVSSKSTSKLWQDFRNKLSILVSTTLILRGVTFPSVDVFIQANHHLPFK 381
 V + + R + VQ FRD +L +L++TTILRGVT P V V+ A +PT+
 Sbjct: 351 GVHCRTASVHAEDKHKEKVVQFRDQQLDLITTLILRGVTVPKVQTVGLGAEISSIFTE 410
 Query: 382 SSLVQISGRVGRALERPEGLLYFLHDGKSKSMHQAIKEIKMNMHI 426
 S+LVQI+GR GR E +G + + H GK+KSM A K IK MN +
 Sbjct: 411 SALVQIAGRTGRHKEYADGVIFYFHFGKTSMLDRAHGIKEKMNEL 455

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3089> which encodes the amino acid sequence <SEQ ID 3090>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.09 Transmembrane 304 - 320 (303 - 322)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2635 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

IGB:U56901 involved in transformation [Bacillus subt... 258 1e-67
 >GP:AA04940 GB:U56901 involved in transformation [Bacillus subtilis]
 Identities = 155/435 (35%), Positives = 249/435 (56%), Gaps = 20/435 (4%)
 Query: 10 RLLLESQLPDSAKQLAQPLK-----SVVILRGKMIQRCYQLDEBA-----RLP83 56

-1116-

		R LL ++L S ++ +K S+ I ++ C RC Q D+
	Sbjct: 22	RHLRLTELSFSDMEIWHIKNGYITAEISINKRKYRCNGC-QTDQRYPSFTHSSGKN 80
5	Query: 57	AYYCRPCLVFRNOSDKLVAIPMHFP--KGNLYWGGQLTAYQEMISQQLLNINNOQK 114
		YCR C++ GR + LY+ + K L W G L++ Q+ + L+ + ++
	Sbjct: 81	KLYCRSCVMGRVSEEVFLYSWKEENSNWKSILKLTWDGKLSGQQKAANVLIEAISKE 140
	Query: 115	TLVHAVTGAGKTEMIYAAITRAVINTGGWCIASPRVDVCVEVATRLSQAAPS-CSICLM 173
		L+ AV GAKTEW+ + IE+ +N G VCI+PR DV +E+A RL AP I ++
10	Sbjct: 141	ELLINAVCGAGKTEMLFPGIESALNGQLVCIATPRTDVVLEAPRLKAAFPQADISALY 200
	Query: 174	AESLPYQR-APIVATTHQLLKPHKAFDILLIDEVDAPFPVNNIQLHYAASONLKEGGAK 232
		S R +P+++TTQHLL++ A D+IIDEVDAFP+ + L +A +A K+
15	Sbjct: 201	GSSDDKGRLSPLMISTTHQLLYKDAIDVMIIDEVDAPFYSADQTLQFAVQKARKNSTL 260
	Query: 233	ILLTATSTRITLERKUNKGEVVKLTARRFHNRPLVIPKPIRSFNLPMIHRQKLPLKILK 292
		+L+AT +L+RK G++ + + R H +PL P+FA N K +R K+P + +
	Sbjct: 261	VYLSATPPKELKRIKALNGQLHSVRIPARHHRKPLPEPRFVWCGNWKKLNRKIPPAVKR 320
20	Query: 293	YLKORKTGYPLLIPLFTIIMAESVTAIKELLPAEQIACVSSQSNRREDITAFRQGGK 352
		+++ K G P+ +P+P++ + E A K + + A V ++ +ARKE + FR G+
	Sbjct: 321	WIEPHVKGRPVFLFVPSVILEKAAACPKGV--HCR TASVHAEDKIRKEVQVPRDQGL 378
25	Query: 353	TILITTSILERGVTFPQIDVFLGSHHRVYSSQSLVQIAGRVGRSIDRDPDTLYPFHGEI 412
		+LITT+ILERGV T P++ VLGA+ +++ +IWQIAGR GR + DG + +PH G
	Sbjct: 379	DLITTTILERGVTVKPVQTVGLGAESSIPTESALVQIAGRGRUIKEVADGDVYIHPGR 438
	Query: 413	SKAMLLARKEIKEMN 427
		+K+ML ARK IKEMN
30	Sbjct: 439	TKSMLDARKHIKEMN 453

An alignment of the GAS and GBS proteins is shown below.

Identities = 223/427 (52%), Positives = 299/427 (69%)

35	Query: 1	MENYLGRNLTKAQLESQLRKIAISLESPIKSGSDYICTRSSVAVNRQGLPTGNFYCRBC 60
		+EN GR L ++QL + ++A L S + + IC RC + + +LP+G YYCR C
	Sbjct: 4	IENSYGRLLLESQLPDSAKQLAQPLKSVVILRGMIQORCHYQLDBEARLPGSAYYCRFC 63
	Query: 61	IYPGRVTSNENLYYFPQKTPSKTNSLAWKGLTFYQNEVSEELLKGISSEKENLVHAVTG 120
		+VEGR S++ LY P F K N L W G L+T YQ +S++LL + +++ LVHAVTG
40	Sbjct: 64	LVPRGNQSDKLYAI PPMHFPKGNLYWGGQLTAYQEMISQQLLNINNOQKTLTVHAVTG 123
	Query: 121	AGKTEMIYHVARVIDTGGSVCIASPRIDVCLLEYKRLSDNPRCAITLMHGBSPYORSP 180
		AGKTEMIY ++ VI+TGG VCIASPR+DVC+E+ RL S F C+I LMM ES YOR+P
45	Sbjct: 124	AGKTEMIYAAITRAVINTGGWCIASPRVDVCVEVATRLSQAAPS-CSICLMHARSFLYQAP 183
	Query: 181	LTIATTHQLLKPHYAFDILLIDEVDAPFPVNDPILYQGVKQALKENGTSLPTATSTTEL 240
		+ +ATTHQLKP+ AFDILI+DEVDAFP+V+N L+ QALKR G I LTATST L
50	Sbjct: 184	IIVATTHQLLKPHKAFDILLIDEVDAPFPVNNIQLHYAASQALKEGCAITLLTATSTRIL 243
	Query: 241	BRKVARKEKLKLHARRFIANPLVPIPMWVWVGSIQSLQTKLPPKLVOLINKRQRTYP 300
		BRKV + E+ KL LARRFH PLVIP+ + + K + OKLP K+ + +KOR+T YP
	Sbjct: 244	BRKVNKGEVVKLTARRFHNRPLVIPKPIRSFNLPMIHRQKLPLKILYKLRQRTGYP 303
55	Query: 301	LLLPFPHHSBQVFTILRQAFPMKEIGFVSSKTSRLKLVQDPDRNKLISLIVSTILR 360
		LL+P P I + T II++ P +I VSS+S R + + FR K +IL+ +T+ILR
	Sbjct: 304	LLIPLFTIIMAESVTAIKELLPAEQIACVSSQSNRREDITAFRQKATILITISILR 363
	Query: 361	GVTFPSVDVFIQANHHLFTKESLWQISGRVGRALERPEGILYFLHDKGSKSMQAIKEI 420
		GVTFP +DVFV+ ++H +++ SLVQI+GRVGR++ +RP+G LYF H+G SK+M A KEI
60	Sbjct: 364	GVTFPQIDVFLGSHHRVYSSQSLVQIAGRVGRSIDRDPDTLYFFHGEISKAMLLARKEI 423
	Query: 421	KNNNHIG 427
		K MN+ G
65	Sbjct: 424	KNNNYKG 430

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1006

A DNA sequence (GBSx1066) was identified in *S. agalactiae* <SEQ ID 3091> which encodes the amino acid sequence <SEQ ID 3092>. This protein is predicted to be comf operon protein 3 (comFC). Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0894 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AA044942 GB:U56901 involved in transformation (Bacillus subtilis)
Identities = 76/230 (33%), Positives = 118/230 (51%), Gaps = 11/230 (4%)

Query: 1 MTCLLCHIEDLSQLT FVELMLLPKQNVICQTCRGSFEALSREMGCOCTCCQ-QIPKQCO 59
M CLLC +T+ L LLKP + V C +C+ ++ + C C + Q C+
Sbjct: 1 MCLLCDSQFSQDV FVRLFLLPKDEKV-CYSRCKLKKITGHI-CPLOGRQPSVHVCV 58

Query: 60 DCIYWGKKGIEV----NHFSLYRYNEAMKKNFSLFKPGQDYLLKDVFTKEIKAAALKCY-- 113
DC W + + S+Y YN+ MK+ S FKP+GD + + F + + K
Sbjct: 59 DCEVNRTRIRDSILLRQNRSVYTTNDMMKTELRSRFPKFGDAEI INAFKDSFSTFSFKVYP 118

Query: 114 -KGYTIVPVPLSHQYQNRQFNQVIAFLQSANIPYKNILSKKDGKQSANNKEERLKQVQ 172
K + +VP+PLS E + R FNQ + + P + L + + KQS K ERL
Sbjct: 119 DKHFWLPIPLSKEREERGFQAHLLABCLDRPSHHPLIRLANNEKSKKKKTRLLSEC 178

Query: 173 QFTLQNEARLGDNLLIVDDIYTTGATIAQIRKLLEBKQ-INKIKSFSLAR 221
F KN + G N++++DD+YTTGAT+ + L EKG ++ SF+L R
Sbjct: 179 IFTDGNNSAGNNIIILLDDLYTTGATLHFAARCLLEKGAASVSSFTLIR 228
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3093> which encodes the amino acid sequence <SEQ ID 3094>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0763 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 100/222 (45%), Positives = 139/222 (62%), Gaps = 2/222 (0%)

Query: 1 MTCLLCHIEDLSQLT FVELMLLPKQNVICQTCRGSFEALSREMGCOCTCCQIPKQCO 60
M CLLC +I ++ ++ E++ L+ + IQC C+ SF+ + + C TCC C+D
Sbjct: 1 MCLLCQGISCTPISITETILFRRISSTPCQQQKSPQKIGKSV-CATCCANSDIACRD 59

Query: 61 CIYWGKKGIEVIRHPSLYRYNEAMKKNFSLFKPGQDYLLKDVFTKEIKAAALKCY-KGYTIV 119
C+ W KG VNH SLV YN AMK PS +KPGQDYLL+ VP E+ + KY KGY V
Sbjct: 60 CLGHWNGXNYVNIHRSILCYNAAMKAYPSQYKPGQDYLLRKVFVAVELADVITTKYKGYIPV 119

Query: 120 VVPLSHQYQNRQFNQVIAFLQSANIPYKNILSKKDGKQSANNKEERLKQVQFTLQNE 179
VVP+S ++ RQFNQV A L++AN+ Y ++ K D QS+ K+ERL + + L
Sbjct: 120 VVPVSPGFRERQFNQVSAILEAANVSYLSLFEKLDNTHQSSKTKERLLVEKSYRLKLV 179
```

Query: 180 AELGENLLIVDDIYTTGATIAQIRKLLSEKGINIKSFSLAR 221
 + + D + LTVDDIYTTG+TI +RK L + +IKS G+AR
 Sbjct: 180 SNIPKELIVDDIYTTGSTIIALRKQLAKVANGSIKSLIAR 221

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1007

A DNA sequence (GBSx1067) was identified in *S.agalactiae* <SEQ ID 3095> which encodes the amino acid sequence <SEQ ID 3096>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3889(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CA91549 GB:A249134 hypothetical protein [Lactococcus lactis]
 Identities = 107/185 (57%), Positives = 140/185 (74%), Gaps = 3/185 (1%)
 Query: 1 MIKYSIRGENIEVTEAIREYVETKLSKVEKYFNEAQELDTRVNLKVYREKTAKEVETILI 60
 MIK+IRGEN+EV+AIR YVE K+ K+K+K+K+ B+ VNLKVY EK AKVEVT+
 Sbjct: 1 MIKFNIRGENIEVTDAIRAYVEDKIGKLDKRYFNDGHEVTAYVNLKVITEKRAKEVETLPA 60
 Query: 61 DSITLRAEDVSQDMYSGIDLVVDKIERQIRKNTKIARKYREKIPASQVPTTEFEAREPDE 120
 ++TLRAED SQDMY SID V +K+K+K+K+K+ K R + +P QVF EF
 Sbjct: 61 KQVTLRAEDVSQDMYSGIDFVEEKLERQIRKYKTRKRNKPRNAVPTQVDFGFEAPLDTT 120
 Query: 121 EAVSQ--RIVRTGNVNLKPMQVEEALQMEILLGHDPFFIYTDADNTNVLVYKREDGELG 177
 + V++ -IVRTK+V LKPMQ EEA+LQM++LGHDF++TDA+ N T+V+Y+R DG G
 Sbjct: 121 DEVAEDHVDIVRTKRVALKPMQAEAVLQMDLGHDPFVPTDADGNTRVYRTRDGRYG 180
 Query: 178 LIEAK 182
 LIE +
 Sbjct: 181 LIETE 185

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3097> which encodes the amino acid sequence <SEQ ID 3098>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3751(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 145/182 (79%), Positives = 165/182 (89%)
 Query: 1 MIKYSIRGENIEVTEAIREYVETKLSKVEKYFNEAQELDTRVNLKVYREKTAKEVETILI 60
 MIK+IRGENIEVTEAIR+YVE+KL+K+K+K+ + QE+D RVNLKVYRE++K+VEVT +
 Sbjct: 1 MIKFSIRGENIEVTEAIRDYVESKLTKEIKYFPAKQEDAIRVNLKVYREKSKVEVTIPL 60
 Query: 61 DSITLRAEDVSQDMYSGIDLVVDKIERQIRKNTKIARKYREKIPASQVPTTEFEAREPDE 120
 DS+TLRAEDVSQDMYSGIDLVVDKIERQIRKNTKIARK+RK+P QVPTTEFEARE
 Sbjct: 61 DSVTLRAEDVSQDMYSGIDLVVDKIERQIRKNTKIARKHREKRVPTQVPTTEFEAREVD 120

-1119-

Query: 121 EAVSQRIVRTKENVLKPMQDVEEALLQMEELGHDFFIYTDADNTINVLVKREDGELGLIE 180
 E ++VRTKENV LKPMQDVEEA LQMEELGHDFFIYTD+ED TN+LY+REDD LGLIE
 Sbjct: 121 EIPEVQVVRTKENVTLKPMQDVEEARLQMEELGHDFFIYTDSEGDATNLYLRREDGELGLIE 180

5 Query: 181 AK 182
 AK
 Sbjct: 181 AK 182

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 10 vaccines or diagnostics.

Example 1008

A DNA sequence (GBSx1068) was identified in *S.agalactiae* <SEQ ID 3099> which encodes the amino
 acid sequence <SEQ ID 3100>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0685 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 20 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 25 vaccines or diagnostics.

Example 1009

A DNA sequence (GBSx1077) was identified in *S.agalactiae* <SEQ ID 3101> which encodes the amino
 acid sequence <SEQ ID 3102> (sgaT). Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -5.95	Transmembrane	99 - 115 (87 - 115)
INTEGRAL	Likelihood = -3.50	Transmembrane	43 - 59 (42 - 60)

30 ----- Final Results -----
 bacterial membrane --- Certainty=0.3378 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 35 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:BAB03942 GB:AP001507 unknown conserved protein [Bacillus halodurans]
 Identities = 47/111 (42%), Positives = 76/111 (68%), Gaps = 5/111 (4%)

Query: 1 MAIIYLIIVAVFAG--EAYIAKEI---SNGVNGLVYALQLAQQAAGVFFVILAGVRLITGE 55
 M I++L+ A+ + A+E+ S + +YA+ + FA G+ V+L GV++ +GE
 45 Sbjct: 233 MGIILFLVGATILALKDITQGAQLIAQSGQSFFPIYAITQSPMFAGGIATVILGVKMFGE 292

Query: 56 IVPAPFKGISSEKLVFNSKPAIDCPVVFYPAFNAVILGFISKVFGELVSMIVM 106
 +VPAF GI+ KLVP ++PALD P+V+P APNAV++GF+ FVG L+ ++V+
 50 Sbjct: 293 VVPAPNGIATKLVPGARPAIDAPVVFMAPNAVILGFLGAFVGLIWLIVTI 343

There is also homology to SEQ ID 516.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1010

A DNA sequence (GBSx1078) was identified in *S. agalactiae* <SEQ ID 3103> which encodes the amino acid sequence <SEQ ID 3104>. This protein is predicted to be tryptophanyl-tRNA synthetase (trpS). Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2156 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC05711 GB:L49336 tryptophanyl-tRNA synthetase [Clostridium
longiporum]
Identities = 225/340 (66%), Positives = 271/340 (79%), Gaps = 3/340 (0%)

Query: 1 MTKPIILTGRDPTGKLIHGIVGSLKNRVLLQNEGSYTLFVPLADQQALTDHAKDPQTIV 60
M K IILTGRDPTGKLIHGIVGSLKNRV LQN G Y P+ +ADQQALTD+A++P+ I
Sbjct: 1 MAKELIILTGRDPTGKLIHGIVGSLKNRVOLNSGIVRSPTIMADQQALTINANPEKIR 60

Query: 61 ESIGNVALDYLAVALGDDPNKSTLPIQSQIPELAEISMYMNLVSARLERNPVTVKETIAQK 120
S+ VALDYLAVALG+DP KST+ +QSQIPEL EL-H++Y+NLV+L+RLERNPTVK EI QK
Sbjct: 61 NSLIEVALDYLAVALGIDPLKSTLIVQSQIPELNELIIMEYLNILVTLSRLERNPTVKELIKQK 120

Query: 121 GFGESIPAGFLVYFVAQAADITAFKANIPLVPGTDQKPMIBQTRSVRSFNHAYNCQVLVE 180
F SIPAGFL+YF+VAQAADITAFKA VPGV DQ PMIEQ REIVRSFN Y +VLVE
Sbjct: 121 NFENSIPAGFLIYFVQAADITAFKATTVFVGEDQLPMIEQAREIVRSFNITYQKVLVE 180

Query: 181 PEGTYPENDAAAGRLPGLDGNNAKSKLNNGIFLADMDTVKKKVMYMDPNHIKVEEPG 240
P+ + P+ GRLPD DG AKMSKS+ N I+LAD+ D +K+KVMYMDPNHIK+ +PG
Sbjct: 181 PKAVIPKG-TIGRLPOTDGNKAKMSKIGNAIYLADRADVIKQKVMYMDPNHIKVTDFG 239

Query: 241 QIEGRNVFHYLVDFGDEDDKEITAMKEHYQKGLGVETKRYLLDILEREISPIREERL 300
Q+BN VF YLD F +D + E MK HY +GGLGVNK K++L +IL+ EL PIR RR
Sbjct: 240 QVGGITVFTYLTDFCKDTETLE--EMKAHYSRGLGVNVKVPFLNELIQALEPIRNRK 297

Query: 301 EYAKDMGQVYQMLQKGSSEKQAQVAASTLDEVKSMAGLNLYF 340
E+ Kd+ +VY++L++GSEKA+ VAA TL EV+ +G+ YF
Sbjct: 298 EFQKIDIPSVYRLKEGSEKAREVAAGTLKEVRETIIGIEYF 337
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3105> which encodes the amino acid sequence <SEQ ID 3106>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2737 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 290/340 (85%), Positives = 316/340 (92%)

Query: 1 MTKPIILTGRDPTGKLIHGIVGSLKNRVLLQNEGSYTLFVPLADQQALTDHAKDPQTIV 60
MTKPIILTGRDPTGKLIHGIVGSLKNRV LQNE Y +FVPLADQQALTDHAK+ + I
```

-1121-

Sbjct: 2 MTKPIILTGDRTYKHLGHVVGSLNRPVFLQENKYMFPVLADQALTDHAKESLIQ 61
 Query: 61 ESIQNVALLNYAVGLDPKSTLFTIQSQIPFLARLSMYNNLVSLARLERNFTVKTIAQK 120
 ESIQNVALLDYL+VGLDP +ST+FTIQSQIPFLARLSMYNNLVSLARLERNFTVKTIAQK
 5 Sbjct: 62 ESIQNVALLDYL+VGLDPKQSTLFTIQSQIPFLARLSMYNNLVSLARLERNFTVKTIAQK 121
 Query: 121 GFQESIPAGFLVYVPAQAADITAFKANLVVPVTDQKPMIEQTREIVRSFNHAYNCQVLE 180
 GFQESIP+GFLVYVPV+QAADITAFKANLVVPV DQKPMIEQTREIVRSFNH Y+ LVE
 10 Sbjct: 122 GFQESIPSGFLVYVPAQAADITAFKANLVVPVNDQKPMIEQTREIVRSFNHTYHTDCLVE 181
 Query: 181 PBGIYPENDAGRLPGLDGNAMKSKSLANGIPTADMDTVKKKVMMSMYTDPNHIIKVEEPG 240
 PBGIYPEN+ AGRLPGLDGNAMKSKSL NGI+I+DD DTV+KKVMSMYTDPNHIIK+B+PG
 Sbjct: 182 PBGIYPENKAGRLPGLDGNAMKSKSLANGIYLSDDADTVKKKVMMSMYTDPNHIIKIEDPG 241
 Query: 241 QIBGNMVFHYLDVPRDEDDQKKTAMKEHYQGGGLGVKTKRYLLDILERELESPTRERRL 300
 QIBGNMVFHYLD+P R EDQ +I AMKEHYQ GGLGVKTKRYLLDILERELE+PIRNRRL
 Sbjct: 242 QIBGNMVFHYLDIPARKEDQADIEAMKEHYQGGGLGVKTKRYLLDILERELEAPTRERRL 301
 Query: 301 EYAKIMGQVYQMIGKSGEKAQVAASTLDEVKSAMG1NYF 340
 EYAKIMG+V+MLQ+GS+KA+ VAA TL EYKSAMG+NYF
 20 Sbjct: 302 EYAKIMGSEYFMQLQSGSQARTVAATLSEVKSAMG1NYF 341

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1011

A DNA sequence (GBSx1079) was identified in *S. galactiae* <SEQ ID 3107> which encodes the amino acid sequence <SEQ ID 3108>. This protein is predicted to be carbamate kinase. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0013 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA04684 GB:A7001330 carbamate kinase [Lactobacillus sakei]
 Identities = 199/311 (63%), Positives = 254/311 (80%), Gaps = 3/311 (0%)
 40 Query: 6 QKIVVALGNNAILSTDASAKAQEALINTSKSLVKLIKEGHVDVITGNGPQVGNLLLQQ 65
 +KIVVALGNNAILSTDASA AQ +A+ T K LV +K+G +I++HNGPQVGNLL+QQ
 Sbjct: 4 RKIVVALGNNAILSTDASANAQIKAVKSTYKQLVAFVKGQDQLIISHNGPQVGNLLTQQ 63
 45 Query: 66 AASDSEKFPAMFLDTCVAMTEGSIGFWLQNALNNELQEQGIDKEVATVTVQIVDERKDA 125
 AASDSEK FAMPFLDT AM++G IG+H+QNA N L E+G+ +VAT+VTQ IVD KD+A
 Sbjct: 64 AASDSEKFPAMFLDTVGAMSGEIGYWMQNAFNEVLAESGLADVATVTVQIVDERKDA 123
 50 Query: 126 FINPTKPIGPFSLSHEDAKKQQAQ-ETGSKFKHEDAGRGWRKVPVSPKVGKIEASVIRLWD 184
 F NPTKPIGPF SE +AKKK + F EDNRGRWR+VVPSP+P+GI+EA VI++LV+
 Sbjct: 124 FQNPTKPIGPFYSHEAAKKQQSINPEAFVFDVADGRGWRVVPSPRPIGQAPVQLKVE 183
 Query: 185 SGVVVISAGGGGVPIEDANTKALGVEAVIDKDFASCTLSELVDADFLVILAGVDERFV 244
 V+ ISAGGGGVPI ++ N I+GVEAVIDKDFAS+ L+ELV AD+ I+LT VDNV+V
 55 Sbjct: 184 GNVTISAGGGGVPAKBN- KLRGVEAVIDKDFASEKLAEVGAADMLLIITAVDNVTV 241
 Query: 245 NFNKPNQEKLERVTVSQMKQYITENQFAGSMLEPKVKAALAFVSNKPSRAITSLLENID 304
 NFNKPNQ+KL V+V++++ YI +QFA GSMLEPK+ AI +V N+P+S+AITSL+N+
 Sbjct: 242 NFNKPDQKKLNVNSVALEEDYIKDDQFAGKSMLEPKIQTAIEYVNNRPDSKAITSLDENVK 301
 60 Query: 305 NVLAQNAQTQI 315

-1122-

N+LA +AGT I
 Sbjct: 302 NLAHDAGTII 312

- A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3109> which encodes the amino acid sequence <SEQ ID 3110>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have no N-terminal signal sequence

- Final Results -----
 bacterial cytoplasm --- Certainty=0.0013 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- Identities = 275/312 (88%), Positives = 295/312 (94%)
- Query: 6 QKIVVALGGNAILSTDSASAKAQREALINTSKSLVKLIKEGHVDVITHGNGPQVGNLLQQ 65
 Sbjct: 4 QKIVVALGGNAILSTDSASAKAQREALI+TSKSLVKLIKEGH+VIVTHGNGPQVGNLLQQ 63
- Query: 66 AADSEKGNFAMPLDTCVAMTEGSGFWMQNALNNELQEQGIDKEVAVVTQVIVDEKQDA 125
 Sbjct: 6 AADSEKGNFAMPLDTCVAMTEGSGFWMQNALNNELQEQGIDKEVAVVTQVIVDEKQDA 123
- Query: 126 FNTPTKPIGPFLTEEDAKKQAEQTSKFKEDAGRGWRKVVPSPKEVGIKEAVIRRLVDS 185
 Sbjct: 124 FNTPTKPIGPFLTEEDAKKQAEQTSKFKEDAGRGWRKVVPSPKEVGIKEAVIRRLVDS 183
- Query: 186 GVVVISAGGGGVVPIEDANTKALKGVEAVIDKDFASQTLSELVDADLFIVLTGVDNVFVN 245
 Sbjct: 184 GVVVISAGGGGVVPIEDANTKALKGVEAVIDKDFASQTLSELVDADLFIVLTGVDNVFVN 243
- Query: 246 FNKPQKLEEVTVSQMKQYITENQFAPGSMPLPKVEAALAFVENKPFESRAITISLENIDN 305
 Sbjct: 244 FNKPQKLEEVTVSQMKQYITENQFAPGSMPLPKVEAALAFVENKPFESRAITISLENIDN 303
- Query: 306 VLAQAGTQIVA 317
 Sbjct: 304 VLSANAGTQIIA 315

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1012

- A DNA sequence (GBSx1080) was identified in *S.agalactiae* <SEQ ID 3111> which encodes the amino acid sequence <SEQ ID 3112>. This protein is predicted to be permease (potE). Analysis of this protein sequence reveals the following:

- Possible site: 52
 >>> Seems to have an uncleavable N-term signal seq
- INTEGRAL Likelihood = -12.63 Transmembrane 450 - 466 (441 - 478)
 INTEGRAL Likelihood = -8.97 Transmembrane 236 - 252 (231 - 259)
 INTEGRAL Likelihood = -8.70 Transmembrane 283 - 299 (277 - 308)
 INTEGRAL Likelihood = -8.44 Transmembrane 165 - 181 (153 - 186)
 INTEGRAL Likelihood = -7.96 Transmembrane 129 - 145 (126 - 151)
 INTEGRAL Likelihood = -6.16 Transmembrane 396 - 412 (394 - 415)
 INTEGRAL Likelihood = -5.15 Transmembrane 45 - 61 (38 - 63)
 INTEGRAL Likelihood = -4.94 Transmembrane 335 - 351 (334 - 352)
 INTEGRAL Likelihood = -3.72 Transmembrane 13 - 29 (10 - 30)
 INTEGRAL Likelihood = -2.92 Transmembrane 417 - 433 (417 - 435)
 INTEGRAL Likelihood = -1.54 Transmembrane 360 - 376 (360 - 376)

-1123-

INTEGRAL Likelihood = -0.53 Transmembrane 207 - 223 (207 - 223)

----- Final Results -----

- 5 bacterial membrane --- Certainty=0.6052(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10295> which encodes amino acid sequence <SEQ ID 10296> was also identified.

- 10 The protein has homology with the following sequences in the GENPEPT database.

>GP_CAA76779 GB:Y17554 permease [Bacillus licheniformis]

Identities = 265/470 (56%), Positives = 347/470 (73%), Gaps = 3/470 (0%)

- 15 Query: 5 MEKEKKLGLLPLMLVIGSLIGGIFDLGMSSRAGLVMLLAWVITAGMOTFLVLSFQ 64
M + EKKLGL L LVIGS+IGG F+L +M+S AG +L+ W+IT +GM SFQ
Sbjct: 1 MAREKKLGLFALIALVIGSMIGGAPNLASDMSAGAGAILIGWITTVGMIALAFSFG 60

- Query: 65 NLSEKRPDLTAGIFPSYAKBFGNFMGPNSSAWGYLGSAMLGNVAYALLFSSLOYFFKFFG 124
NL+ KRPLD GIF+YA+EGF+FMGPNSS WGYW +A LNVAY LLFS+GYF FG
20 Sbjct: 61 NLTKRPLDGGIFTYAKBFGFMGPNSSAWGYWFAALLGNVAYGTLFSAIGVTFIPAG 120

- Query: 125 NGNNIISIIIGASIVHVVHFLIRGVNIAAFINTIVTAKLVFVPIFLISALLAFKFNIF 184
+G NI SIIGAS+++W VHFILIRGV +AA IN I T +KLNV+ F+I+ + F ++F
Sbjct: 121 DQNIASIIIGASVILWCVHFLIRGVQSAAMINLITISKLVFIFAFILAIIFVPHDLF 180

- 25 Query: 185 SLIDWGNLH-QSIFNQNSTMTKTAVMVPIEGAVVFSRAKKHSDIGKASILAFMI 243
+ D WG GL SI QV STM VWF GIEGAV+PS RAKK SD+GKA+++ L +++
Sbjct: 181 TNDPWGKGLSLGIGTGVKSTMLTVVWPTGIBAVLPSRAKSSDVGKATVIGLISVL 240

- 30 Query: 244 SLVVLISVLISGIMSRPELANIKTPMAYVLRKAVHMGAILVNLGVII SVPGAILAWTL 303
+YV+I++LSLG+M++ Ld L P+MA ++E VG WGA+L+NLG+IISV GA LAWTL
Sbjct: 241 VIYVMITMLSLGVNQNLAEELNPNMAIMEHIVGKGAVLINLGLIISVIGAWLAWTL 300

- 35 Query: 304 FAELPYQAAKEGAPKFFAKENKKNKAPINSLLVNLVQAFILITFLPTQSAVRFQFALA 363
FA ELP AA+EG FFK+F KENK AP N+L +TN +Q FL+TFL + +AY+F F+LA
Sbjct: 301 FAGELPLIAAREGVFPKWFGENKKNKAPTALNLITNAIQELFLITFLISDAVYQAFELA 360

- Query: 364 SSAILIPYAPFALYQQLPTLRDQSTFQHKNNLIIGILATIVAVVLYAGGFVYLLLMI 423
SSAILIPY F LYQL+++ + P KNLIIIGI+A+IY V+L+YA G DYLLLMI
40 Sbjct: 361 SSAILIPYLSGLYQLKYSWLHKE--PNRKNLIIGIISIVGWLVAAGLYLLLMI 418

- Query: 424 AYTLGMILYIKMKDKKLPFVGVYKESIAIVALLCLLCIEIMTQDIDI 473
Y G+++ +RK + P+F E + A +IL L ++ +I + +G I I
45 Sbjct: 419 LYAPGLIVFRAVRKGRQFVFNKARELLAALILVLAIVRIASGSGISI 468

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3113> which encodes the amino acid sequence <SEQ ID 3114>. Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have no N-terminal signal sequence

- 50 INTEGRAL Likelihood = -11.52 Transmembrane 331 - 347 (327 - 354)
INTEGRAL Likelihood = -9.50 Transmembrane 390 - 406 (383 - 410)
INTEGRAL Likelihood = -8.12 Transmembrane 50 - 66 (45 - 75)
INTEGRAL Likelihood = -7.59 Transmembrane 235 - 251 (234 - 262)
INTEGRAL Likelihood = -6.21 Transmembrane 133 - 149 (128 - 151)
55 INTEGRAL Likelihood = -5.84 Transmembrane 162 - 178 (153 - 183)
INTEGRAL Likelihood = -2.02 Transmembrane 105 - 121 (105 - 121)
INTEGRAL Likelihood = -1.49 Transmembrane 414 - 430 (414 - 431)
INTEGRAL Likelihood = -0.69 Transmembrane 280 - 296 (280 - 296)
INTEGRAL Likelihood = -0.59 Transmembrane 21 - 37 (21 - 37)
60 INTEGRAL Likelihood = -0.32 Transmembrane 205 - 221 (205 - 222)

----- Final Results -----

bacterial membrane --- Certainty=0.5607(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP:AAB85052 GB:AB000837 cationic amino acid transporter related
protein [Methanobacterium thermoautotrophicum]
Identities = 108/422 (25%), Positives = 213/422 (49%), Gaps = 36/422 (8%)

10 Query: 26 INAVIGSGIFLLPRAIYKGLGPASIAVMFGTALLTIMLAVCFPAEVSQYPGKNGAPQYSK 85
+ ++G+ I+++ LGDASI ++ +AA+ F+E S + G3 + Y+
Sbjct: 19 VGTIAGADIYIVAAVAGSGLPASILAWLIALGLMALIILVSEASMLPRTGPPYVYAG 76

Query: 86 RAPGDFIGFNVGFLGNTVTIPAWAAMAGFARMFIITPAEGWHIFL-----SIGL 137
A G F GF GN++ +W A+A +F + F + + IDL +
Sbjct: 79 BALGRFTGF---ITGNSLNVSSWVAIA----VFPLAFIYYLYSTPLDPPAAVAVIKVLF 130

Query: 138 IILLSLMNIAGLNTSKIVTITATIKLIPVAFPCACTLPIFKNG---LPMFTPPVLEP 193
I+ L++NIAG+ + V TI K+ P+ F + + N+TP +
Sbjct: 131 ILSLTIINAGVGRAGKNDILITLKVAVLLFAVLGAHLALNGLLVSNYTPAEPFG- 189

20 Query: 194 GTNLIGASINTAVYIFYGFGFETLSIVAGEMRDPEKIVFRALLGSISVSLVNLIGG 253
LGA+ V +F+ ++GFE +++ A E+RDPE+ P ++ + V+ Y+L
Sbjct: 190 ----LGAIGTIVTVLFWAYVGFELVTVVEADEVRDPTPLSLITGMIFVTLFYILTNV 245

25 Query: 254 TIAMLSQIMNTN-APVQDAFVKMIGCAPANVVSIGALISITGLMGESIMVRYGAIA 312
++ ++ ++ ++ AP+ A ++G GA ++ GA+ SI G + R A++
Sbjct: 246 ILGLVPRVRLASTAFLTVAGYSLMGGIGALITAGAVFSIAGSEAGMLTARLLFMS 305

30 Query: 313 DEGLLPAALAKQNN-GAPLVAIVLVSAGATAIVLLTGSFSLAKLSVFRFPYIPITAL 371
++G LP +++ ++ G P ++ILV A++ LTG+ L +SVV Y T++
Sbjct: 306 EDGLFPGFLSRVRRRFGTPHMSILVQNLITALLAALTGVSGLLESVTLTLVYAVTCIS 365

Query: 372 VMKLRKDDPDANVIFRPVPGPIIPILAVIVSLVMVINGDPMNPFVYAGVVIASSVYYLM 431
+ LR+ D P+ +L V+V ++ P +G +I++ + YL+
35 Sbjct: 366 LAILRRRDSGI-----PLKSVLGLVLCVILMNTPTSTTANGLL-LILSGAPLYLI 416

Query: 432 HG 433
G
Sbjct: 417 PG 418

An alignment of the GAS and GBS proteins is shown below.

Identities = 104/368 (28%), Positives = 162/368 (43%), Gaps = 32/368 (8%)

45 Query: 1 MRYTKEKKEKLLGLPLTLMVIGSLIGOGIFDLMMNMSRAGLVPLMLIAWVI-TAIGNTGF 59
M + ++ K L T+ I ++IG GIF L ++ GL P IA + TAI
Sbjct: 6 MNEQERBQAKFSLGNTLYGINAVIGSGIFLLPRAIYK--GLGPASIAVMFGTALLTIML 63

Query: 60 VLSFQNLGEBKRPDLTAGIFSYAKEGFGNFMFGNSA--WGYWLSAMLGNAVYALLFSSL 116
+ F +S G F Y+K FG+P+GFN W + AM A A +F
50 Sbjct: 64 AVCTFAEVSQYFGK-NGGAPQYSKRAFQDFIGFNVGFLGNTVTIPAWAAMAGFARMFIIT 122

Query: 117 GYFFKFPFGNNIISIIIGASIVIMVHFLILRGVNTAIPINTIVTEAKLVFVLIPLSAL 176
F+ G+I IG I++ ++ + G+ T+ + T AKL+P++ P L
Sbjct: 123 FPAFE-----GWHIFLSIGLIIILSLAN---IAGLKTSKIVTITATIAKLIPIVAFPCACTL 175

55 Query: 177 LAFK-----FNIPSLDIWINGLHQSIFNQVNSMTKTAUVVFIGIEGAVVFSGRACKHSDI 231
K F F G L +I N TAV++F G F S A + D
Sbjct: 176 FFIKNGLMNFTFPVQLEGTINLLGAIN-----TAVIYFGFIFGFTLSIVAGENRD 228

60 Query: 232 GKASILALPIMISLYVLISVLSLG---IMSRPELANAKTPAM-AYVLEKAVGHMGAFLVN 287
K AL IS+ ++ +L +G M ++ P A+V K +G GA +V+
Sbjct: 229 EKNVFRALLGISISIVSLYMLIIIGTITAMIGSQIMNTNAPVQDAFV--KMIGFAGAKRVVS 286

Query: 288 LGVILSVFGAILAWTLFAELPYQAAKEGAPKPFKAKNNKAPINSLLVNLICVQAFLI 347
+G +IS+ G + ++ A EG P AK+N+N AP+ ++LV+ L+
65 Sbjct: 287 IGALISITGLNMGESIMVPRYGAIADEGLLPAALAKQNNKAPLVAIVLSGATAIVLLL 346

-1125-

Query: 348 TFLFTQSA 355
 T F A
 Sbjct: 347 TGSFSLA 354

5

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9079> which encodes the amino acid sequence <SEQ ID 9080>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

10	INTEGRAL	Likelihood = -9.92	Transmembrane	77 - 93 (72 - 100)
	INTEGRAL	Likelihood = -9.29	Transmembrane	279 - 295 (274 - 303)
	INTEGRAL	Likelihood = -9.08	Transmembrane	203 - 219 (199 - 225)
	INTEGRAL	Likelihood = -8.55	Transmembrane	174 - 190 (171 - 197)
	INTEGRAL	Likelihood = -8.33	Transmembrane	436 - 452 (432 - 455)
15	INTEGRAL	Likelihood = -7.32	Transmembrane	329 - 345 (324 - 350)
	INTEGRAL	Likelihood = -5.63	Transmembrane	402 - 418 (396 - 421)
	INTEGRAL	Likelihood = -4.88	Transmembrane	460 - 476 (456 - 479)
	INTEGRAL	Likelihood = -4.51	Transmembrane	379 - 395 (377 - 401)
	INTEGRAL	Likelihood = -2.81	Transmembrane	48 - 64 (48 - 65)
20	INTEGRAL	Likelihood = -2.23	Transmembrane	243 - 259 (243 - 262)
	INTEGRAL	Likelihood = -0.37	Transmembrane	123 - 139 (123 - 139)

----- Final Results -----
 bacterial membrane --- Certainty=0.4970 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

25

An alignment of the GAS and GBS sequences follows:

Score = 62.1 bits (148), Expect = 2e-11
 Identities = 59/250 (23%), Positives = 107/250 (42%), Gaps = 12/250 (4%)

Query: 143 WGSYLKGLLAN--YNIIVLPRALNQTFLN--KNSTYIDILPV-LVMFFVTGIVLMSKLLAL 197
 WG +L L N Y +L ++L F I I + +V+ V ++L A
 Sbjct: 95 WGYLWLANLGNVAIALLFSLGTFKFPGNINLIISITGASIVTVVHFLILRGVNTAA 154

35

Query: 198 RFNSFLVILKFSALALFIFVGIFPDHNMWSHFAPGVGQITGGKTFAGASVMPFAFL 257
 N+ + K + +F+ + N +S +G G + + +F+
 Sbjct: 155 FINTIVTFAKLVFVILFLISALLAFKFNIPS-LDIWGNLHQSIFNQVNSTMTKTAWVVF 213

40

Query: 258 GFESISMAVDEVKEPKQTKPKGILSLIIVTALYIVVTTILTGIV--HYTKLVNPDVA 314
 G E + + K + I K IL+L + +LY++++ + GI+ L P A+A
 Sbjct: 214 GIEGAVVFGRAK-KHSDIGKASTILALFTMISLVLSVLSIGIMSRPELANLKT-PANA 271

45

Query: 315 PALNRILRYWAADYVSTIVALLTLTVCSIMTYALARTIYSISRDLGLPKSLYTLTKKHKV 374
 + L +W A V++ I+++ ++ T A Y ++G PK + RHK
 Sbjct: 272 YVLEKAVGHGAILVNLGVIIISVFGAILAWTLFAELIPYAAKKGAPK-PPAKENKHA 330

Query: 375 PQNATVATKL 384
 P N+ LVT L
 Sbjct: 331 PINSLLVTML 340

50

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1013

55 A DNA sequence (GBSx1081) was identified in *S.agalactiae* <SEQ ID 3115> which encodes the amino acid sequence <SEQ ID 3116>. This protein is predicted to be unnamed protein product (argF). Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

-1126-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3757 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3117> which encodes the amino acid sequence <SEQ ID 3118>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.48 Transmembrane 171 - 187 (171 - 188)

----- Final Results -----

bacterial membrane --- Certainty=0.1192 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP: CAB12563 GB: Z99108 similar to metabolite transporter [Bacillus subtilis]
 Identities = 190/467 (40%), Positives = 284/467 (60%), Gaps = 13/467 (2%)

Query: 25 TIFRKK-----KKYSKTKEMQRHKVIDLVPGLGSMVGTGIFTVTIGAAKYAGPALTI 79
 ++FRKK S + R DL LG+G +GTCUIF +TG AA AGPAL I
 Sbjct: 3 SLFRKKPLETSLAQSKSKSLARTLSAPDLTLGLGICVIGTGTIPVITGTVAATGAGPALII 62

Query: 80 SIITSLAIGILALFYAEFASFMPSNMGAYSYYVATLGEFFAWLVGMVYIMEFLTALSSV 139
 S I++ +A + A YAEF+S +P +G YSY Y TLGE A+L+GM +++E++ A+S+V
 Sbjct: 63 SPFLALACALAAFCYAEFSSSIPISGSYSYSYVTLGELLAPLIGDMLLEYVIALSAF 122

Query: 140 ATGWSYSLKGLLANYNIVLPNALNKTGYTIDILFVLV/MFPVVTGIVLMNKKLALRF 199
 A GM SY + LLA +N+ +P AL G G ++ +++ +T IV K + RF
 Sbjct: 123 ATGWSYFQSLLAGFNHILPAALTGARSGMAGVFNLPAAVILLITATVSRGVKSTRF 182

Query: 200 NSFLVILKFSALALFIPVGIFFIDHNNWSHFAPYGVGQITGGCTGIPAGASVMFFALGF 259
 N+ +V++K + + LFI VGI ++ +NMS F P+G+ G+ A+ +FFA+LGF
 Sbjct: 183 NNIVIVLMKLAILLFIIVGIGYKPDNMSFPMPFGM-----EGVLSAATVFFAYLGF 235

Query: 260 ESISMAVDEVKEPQRTIPKGIILSLIIVTALYIVVTTILTIGIVHYTKLVNPDVAVALRN 319
 +++S A +EVK POK +P GII +L + T LYI V+ +LTTG++ Y KLVN D V+FA+L
 Sbjct: 236 DAVSNASEEVKNPQKNMPVGIISALAVCTVLYIAVSLNLTGMMPYAKLVGDPVSLFKF 295

Query: 320 IRLYWAADYVSIVAILTLITVICISMTYALARTIYSISRDGLLPKSLYTLTKKNNVPQNA 379
 + A +S+ AI+ + TV +++ YA R +++SRDGLLP + K P T
 Sbjct: 296 VGCDAVAGIISVGAIIGITVLMALLYAQVRLTFAMSRDGLLPGLPAKVHPSFTPTFRNT 355

Query: 380 LVTHGLAMICAGIFPLSSIAEFVNICTLAYLILSGAIKKLRRIEGEPKANEKTPLPVF 439
 +TG++A AG L +LA VN+ TLA ++S A+I LR+ E KA+ F+ P VP
 Sbjct: 356 WLVTGIVAAGIAGFINLGTLAHLVNMGTLAAPTVISIAVIVLRKKHPEIKAS-FRVPPFV 414

Query: 440 LPMALIIICLSFMSQYKAPTWIAFAITATIGTLIYLAGYTHSIENK 486
 +P+++ ICL FM TW+P I +GTL+Y Y HS+ NK
 Sbjct: 415 VPIISAGICLWFMYSLPGVTWLSFVITWIAVGLTVYFLYSRKHSLINK 461

An alignment of the GAS and GBS proteins is shown below.

Identities = 312/337 (92%), Positives = 324/337 (95%)

Query: 1 MTQVFQQRSLAEKDFSRREEFYILDFSAHLKDLKRGVPHHYLGGKNIALLEFKTSRT 60
 MTQVFQQRSLAEKDFR E EYLIDFSAHLKDLKRGVPHHYLGGKNIALLEFKTSRT
 Sbjct: 1 MTQVFQQRSLAEKDFTRAELEYILDFSAHLKDLKRGVPHHYLGGKNIALLEFKTSRT 60

Query: 61 RAAFTTAIDLGHANPHYLGANDIQLGKKESTEDTAKVLGRMFDGIEFRGFSQRMVEELAE 120
 RAAFTTAIDLGHANPHYLGANDIQLGKKESTEDTAKVLGRMFDGIEFRGFSQRMVEELAE
 Sbjct: 61 RAAFTTAIDLGHANPHYLGANDIQLGKKESTEDTAKVLGRMFDGIEFRGFSQRMVEELAE 120

-1127-

Query: 121 PSQVPMNGLTDEMHTQMLADYLTIKENPGKLEGLTLYVCGDGRNNVANSLLVAGTLMG 180
 PSQVPMNGLTDEMHTQMLADY T+KENPGKLEGLTLYVCGDGRNNVANSLLV G ++G
 Sbjct: 121 PSQVPMNGLTDEMHTQMLADYFTVKENPGKLEGLTLYVCGDGRNNVANSLLVGTAILG 180

Query: 181 VNVHIFSPKELPFAEIRIVKLAEEYAKESGAHVLTNDVDEAVKGADVFTYDVVWSMGEED 240
 VNVHIFSPKELFP ERIV LAR YAKESGA +L+T++ DEAVKGADV YTDVWWSMGEED 240
 Sbjct: 181 VNVHIFSPKELPFEIRIVTLAEYAKESGARILITDEADEAVKGADVLTVDVWWSMGEED 240

Query: 241 KFERVELLPQYQVNMELIKANNNDLIFLHCLFAPHDINTVYGKVAKKGVKMEVTD 300
 KFERVELLPQYQVNM+L++KA MD LIFLHCLFAPHDINTVYGKVAKKGVKMEVTD 300
 Sbjct: 241 KFERVELLPQYQVNM+LQKAGNDKLI FLHCLFAPHDINTVYGKVAKKGVKMEVTD 300

Query: 301 EVFRSKYARHFDQAENRMHTIKAVMAATLGNLFIPKV 337
 EVFRSKYARHFDQAENRMHTIKAVMAATLGNLFIPKV
 Sbjct: 301 EVFRSKYARHFDQAENRMHTIKAVMAATLGNLFIPKV 337

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

20 Example 1014

A DNA sequence (GBSx1082) was identified in *S.agalactiae* <SEQ ID 3119> which encodes the amino acid sequence <SEQ ID 3120>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0456(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10921> which encodes amino acid sequence <SEQ ID 10922> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3121> which encodes the amino acid sequence <SEQ ID 3122>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -5.41 Transmembrane 121 - 137 (118 - 140)

----- Final Results -----
 bacterial membrane --- Certainty=0.3166(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 An alignment of the GAS and GBS proteins is shown below.

Identities = 65/113 (57%), Positives = 83/113 (72%)

Query: 31 MEEFDDNDDEQDTIYAVLYDGKQPVSTGRFLPETQTEARLIRIATLKGYGRNGYGTAKIII 90
 M ++FD NDE T+YAV+YD QPVSTG+FL ET+ EARLIRI TL Y G GYG K+
 Sbjct: 1 MADKFDANDERTIRYAVVYDNDQPVSTGQFLAETKIEARLIRIVTLADYCGCGYGRVTE 60

Query: 91 ALENYACENGHYHLTIHAEITAKDFYQTLGVQATGNIYMEDEGAQCTLEKYLI 143
 ALE Y + G++ LTIH+ELTA+ FY+ LSYQ+ G +HDEG CQ+L K ++
 Sbjct: 61 ALETYTRREGFYQITIHSELTATQTFYRNILGVQSYGPKLEDGRYCQSLAKTIL 113

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1015

A DNA sequence (GBSx1083) was identified in *S. agalactiae* <SEQ ID 3123> which encodes the amino acid sequence <SEQ ID 3124>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 58
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.2160(Affirmative) < succ>
10  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3125> which encodes the amino acid sequence <SEQ ID 3126>. Analysis of this protein sequence reveals the following:

```

15  Possible site: 58
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.2730(Affirmative) < succ>
20  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

   Identities = 375/411 (91%), Positives = 395/411 (95%), Gaps = 1/411 (0%)

25  Query: 1  MTQTHPIRVFSEIGLKKVMLHRPGKSIENLMPDYLERLAFDDIPFLDAQKHDAPQA 60
   MT  PIRV+SEIGLKKV+LHRPGKSIENLMPDYLERLLFDDIPFLDAQKHDAPQA
   Sbjct: 1  MTAQTPIRVYSEIGLKKVLLHRPGKSIENLMPDYLERLLFDDIPFLDAQKHDAPQA 60

30  Query: 61  LRNEGVEVLYLENLAESLTNQEIREQFIDYIGEANVRGRATKKAIRELLINKNKEL 120
   LR+EG+EVLYLE LAESL  EIRE FIDY+  EAN+RGRATKKAIRELL+ I+DN+EL
   Sbjct: 61  LRDEGIEVLYLETLAESLVTPEIREAFIDEXLSEANIRGRATKKAIRELLMAEDNQSL 120

   Query: 121  IEKTMAGIQKSELPEIPSEKGLTDLVESNYFFAIDPMNPLYTRDPFATIGNVSLNEM 180
   IEKTMAG+QKSELPEIP+SEKGLTDLVESNYFFAIDPMNPLYTRDPFATIG  GVSLENEM
35  Sbjct: 121  IEKTMAGVQKSELPEIPASEKGLTDLVESNYFFAIDPMNPLYTRDPFATIGTVSLNEM 180

   Query: 181  PSETRNRETLGKYIPTHHPYSG3-KVPMVYREBETTRIGGDELVLKDVLAIGISQRT 239
   PSETRNRETLGKYIPTHHP YG3 KVPMVY+R  ETRTRIGGDELVLKDVLAIGISQRT
40  Sbjct: 181  PSETRNRETLGKYIPTHHPYGGGKVPVYDRNETTRIGGDELVLKDVLAIGISQRT 240

   Query: 240  DAASIEKLLVNIFFKQNLGFKKVLAFEFANNRKFHMLDTVFTMVDYDKFTIHEIGDLRV 299
   DAASIEKLLVNIFFKQNLGFKKVLAFEFANNRKFHMLDTVFTMVDYDKFTIHEIGDLRV
45  Sbjct: 241  DAASIEKLLVNIFFKQNLGFKKVLAFEFANNRKFHMLDTVFTMVDYDKFTIHEIGDLRV 300

   Query: 300  YSVTYENQDLHIEEKGLDLADLLAKMLGVEKVLIRCCGNDLVAAGREQQNDGSGNTLTIA 359
   YSVTY+N+LH I  EEKGDLA+LLA NLGVEKV+LIRCCGNDLVAAGREQQNDGSGNTLTIA
50  Sbjct: 301  YSVTYDNEELHIEEKGLDLAELLANLGVEKVLIRCCGNDLVAAGREQQNDGSGNTLTIA 360

   Query: 360  PGVVIVYNRNTITTALESKGLKLIKINGSRLVRGKGGRCMSPFFEREDL 410
   PGVV+VYNRNTITTALESKGLKLIK+GSELVRGKGGRCMSPFFERED+
55  Sbjct: 361  PGVVIVYNRNTITTALESKGLKLIKINGSRLVRGKGGRCMSPFFERED 411

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1016

A DNA sequence (GBSx1084) was identified in *S. agalactiae* <SEQ ID 3127> which encodes the amino acid sequence <SEQ ID 3128>. Analysis of this protein sequence reveals the following:

Possible site: 20

-1129-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 5 bacterial cytoplasm --- Certainty=0.3162 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8703> which encodes amino acid sequence <SEQ ID 8704> was also identified. This protein has an RGD motif and has homology with the following sequences in the GENPEPT database.

- >GP:AA07568 GB:AE004834 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 42/132 (31%), Positives = 74/132 (55%), Gaps = 3/132 (2%)
- 15 Query: 35 IQTYRKAYQTFKTK-KGARSSIRALLKRVNSGNEITSINPLVDIYNAASLRFLPGAED 93
 + + +A++ F K + S EAL KR + SI+P+VD+YNA S++F +P+G E+
 Sbjct: 63 LAAWAFAFRFQAKFQPTCSAEALKKRALRDXGLPSIDPVVDLYNAISVQFAIPVGGEN 122
- Query: 94 SDTFRGDLKLTITTINGDGFYLI--GEDFNRPTLSGELAYVDVGAVCRCPNWRDGKRTMI 151
 + G +L +G + F + GE + GE+ + DD+G CR +NWR G RT +
 20 Sbjct: 123 LAAYAGPPRLVADGSETFDITLNGEALDESDDPGEVVRDGLVTCRRNWRQGVRTSL 182
- Query: 152 TDNTQNAFLVIE 163
 + + + ++E
 25 Sbjct: 183 DASARRMWFIE 194

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3129> which encodes the amino acid sequence <SEQ ID 3130>. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 30 bacterial cytoplasm --- Certainty=0.0700 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
- 35 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 127/199 (63%), Positives = 155/199 (77%)

- 40 Query: 8 ELKQLSDSHSLAKKYLBQKBFQNRVIQTYRKAYQTFKTKKGARSSIEALLKRVNSGNE 67
 +KQLL+DSH LAK YL FS N+V+Q YRKAYQ FKTKKGARSSIEALLKRV++G
 Sbjct: 36 DVKQLLADSHELAKAYLTADNFDSNQQVQVYRKAYQHFKTKKGARSSIEALLKRVNSQGS 95
- Query: 68 ITSINPLVDIYNAASLRFLPGAEDSDTFRGDLKLTITTINGDGFYLIIGEDFNRPTLSGE 127
 I SINPLVDIYNAASLRFLGP GAEDSD+ F GDL+LTIT+GGD+FYLI+ N PTL E
 45 Sbjct: 96 IPSINPLVDIYNAASLRFLPGAEDSDSFIGDLRLTITDGGDFFYLIIGDANNPTLPNE 155
- Query: 128 LAYVDVGAVCRCPNWRDGKRTMITDNTQNAFLVIELIDNGREIIFKEALDFIATVNR 187
 L Y DD+GA CRC NWRDG+RTM+T++T+NAFL+IE +D + +EAL FI + +
 50 Sbjct: 156 LCYKDDIGAPRCINWRDGEKRTMTVETRTKNAFLIEALDOEGQNRQLQALKFIEGSAKMY 215
- Query: 188 LKAKTQTIIIDKEHSRITL 206
 L A T +LDK++ + L
 Sbjct: 216 LHAITSVHVLDKDNPHVPL 234

- 55 SEQ ID 8704 (GBS298) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 2; MW 29kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 48 (lane 5; MW 54kDa).

The GBS298-GST fusion product was purified (Figure 203, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 297), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1017

A DNA sequence (GBSx1085) was identified in *S.agalactiae* <SEQ ID 3131> which encodes the amino acid sequence <SEQ ID 3132>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3770 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1018

A DNA sequence (GBSx1086) was identified in *S.agalactiae* <SEQ ID 3133> which encodes the amino acid sequence <SEQ ID 3134>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4263 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP: CAB95946 GB: Y17554 Crp/Fnr family protein [Bacillus
licheniformis]
Identities = 85/214 (39%), Positives = 126/214 (58%), Gaps = 14/214 (6%)

Query: 11 RQLDDFKHFTIEQPDHIVSHIKHRTALKNHTLFFGQDYREKLFLLIQSGHIVKIQSDQSDAGS 70
R L+D K F I R+ K L F E D R E++L+ G +R+B+S+ +GS
Sbjct: 22 RDL EDMKQF-----IYWRSYHKGQILFMEDDPFRMYLLLDGFIKLEKSNRAGS 70

Query: 71 FIYTDYVRQGTVPFPGGLFLDDDYHPSAVAITDIEYFSLPMALYBEYSLSQINQMHLR 130
YTDYVR T+FP+GGLF D+ YH+AA A+TDIE + +PM ++B+ N N + +
Sbjct: 71 MFYTDYVRPHTLFPFGGLFDEHYHAAALTDIKLYYIPMNI FDLVRDNKNLLYDILN 130

Query: 131 KYSKILRVHEIRLNBMVTSSASMRVQSLATL---LLQVPTRGHLLFPPTITTEIANMSG 187
S +L +HS RL+ + S A RV Q++ L L Q + + PIT RIA +SG
Sbjct: 131 HLSDLALHEERLKRITLSAHDRVTAQIYVLTESLGQKESNSTVINCPITTAETAKISG 190

Query: 188 T*RETIVSHVLKELRQKDIVEMGKKLLYNKNYF 221
T*RETIVS VLK+LR + ++ K++ N YF
Sbjct: 191 TSRETIVSAVLKLRCEGVISQMNKQIMINRPEYF 224
```

-1131-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3135> which encodes the amino acid sequence <SEQ ID 3136>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4478 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 130/224 (58%), Positives = 180/224 (80%)

```

Query: 1  MITKEQYFYPRQLDPFKHFTIEQFDHIVSHIKHRTALKNHTLFFEGDYREKLFQISGHV 60
      +I +E Y Y R+L+DF++F+IEQFD IV ++ R A K+H LFFEGD R+KLF+ SG+
Sbjct: 1  VIRREDYQYLRKLANDFRY78IEQFDKIVGQMEFRKAKDHILFFEGDKRKLFLVTSSYP 60

Query: 61  KIEQSDASGSGFIYTDYVRQSTVFFYQGLFLDDDDYHFAVAITLIEYFSLPMALYEVSLQ 120
      K+EQSD SG+F+YTD+-R GT+FFYQGLF DD YHFS VA+TD+ YF P+ L+E+YSL+
Sbjct: 61  KVEQSDQSGTFMYTDFIRHGTIFPYQGLFTDDYHFSVAMTDVTYFYFPVDLFEDYSLE 120

Query: 121 NINQMCHLCRKYSKLLRVHEIRLNNMTSSASMRVQISLTLQLQVPTERGHLPFFITTI 180
      N QMCHL K SKLL +HE+R+RN++TSSAS RVQISLA LL+++ + LPF +TT
Sbjct: 121 NRIQMCHLYSKMSKILLEHRLKRVNLTSSASRVQISLAILLVEGMSKDSITLFPQLTIT 180

Query: 181 EIANMGTTTRETIVSHVLKELRQKDIVEMKGGKLLNNNNYFKKF 224
      +IA +SGTTTRETIVSHVL++L+++++ +KGL L Y +K+YF ++
Sbjct: 181 DIAIGTTTRETIVSHVLEDLKKQELITIRGKYLTLYLDKDYFLQY 224

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1019

A DNA sequence (GBSx1087) was identified in *S.agalactiae* <SEQ ID 3137> which encodes the amino acid sequence <SEQ ID 3138>. Analysis of this protein sequence reveals the following:

```

Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1643 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2161> which encodes the amino acid sequence <SEQ ID 2162>. Analysis of this protein sequence reveals the following:

```

Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1201 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 461/493 (93%), Positives = 478/493 (96%)

```

Query: 2  NSKWDITKFLKKGFTFDVLLIIPASHVLPNEVDMKTELADNELTNIPIITAMDTVDSK 61
      NSKWDITKFLKKG+TFDVLIIIPASHVLPNEVD+KTELADNELTNIPIITAMDTVT SK

```

-1132-

- Sbjct: 1 MSNWDTKFLKGGTYFDVLLIPAESHVLPNEVDLTKLADNLTLNIPITTAAMDTVTGSK 60
- Query: 62 MAIAIARAGGLGIITHQMSIVDQAEVRVKRSNGVVIDPPFLITEDNTVSEAEELMNY 121
MAIAIARAGGLGIITHQMSI +QAEVRVKRSNGVVIDPPFLTP+ VSEAEELM Q Y
- 5 Sbjct: 61 MAIAIARAGGLGVHKNMSITQAEVRVKRSNGVVIDPPFLITPEHKVSEAEELMQRY 120
- Query: 122 RISGVPIVETLENKRLGVITNRDMRPFISDYKOLISEHMTSQNLVAPITGDLTAERIL 181
RISGVPIVETL NRKLGVITNRDMRPFISDY ISEHMTS++LVTA +GTDLETAERIL
- 10 Sbjct: 121 RISGVPIVETLANKRLGVITNRDMRPFISDYNAPISEHMTSEHLVDAVGTDLETAERIL 180
- Query: 182 HEHRIEKLPLVDDBSGLITTKDIEKVEIEPKAKDEPRGLVAVAGVTSDFPRAE 241
HEHRIEKLPLVDA+ GRLSGLITTKDIEKVEIEP AKDEPRGLLVA AVGVTSDFPRAE
- 15 Sbjct: 181 HEHRIEKLPLVDSGRLSGLITTKDIEKVEIEPFAAKDEPRGLLVA AVGVTSDFPRAE 240
- Query: 242 ALFEGAGDAIVIDTANGHSAGVLRKIAEIRAHFNRMTLIAGNIATNEGARALYDAGVDV 301
ALFEGAGDAIVIDTANKHSAGVLRKIAEIRAHFNRMTLIAGNIATNEGARALYDAGVDV
- 20 Sbjct: 241 ALFEGAGDAIVIDTANGHSAGVLRKIAEIRAHFNRMTLIAGNIATNEGARALYDAGVDV 300
- Query: 302 KVGIGPGSICCTKVVAGVGVPQITAIYDAAVAREYKGTIIADGGIKYSGDIVKALAAG 361
KVGIGPGSICCTKVVAGVGVPQ+TAIYDAAVAREYKGTIIADGGIKYSGDIVKALAAG
- 25 Sbjct: 301 KVGIGPGSICCTKVVAGVGVPQVTAIYDAAVAREYKGTIIADGGIKYSGDIVKALAAG 360
- Query: 362 NAVMLGSMFAGIDEAPGETEIQQRKPKTYRGM3IAAMKKGSSDRYPQGSVNEAMKLPV 421
NAVMLGSMFAGIDEAPGETTI+QQRKPKTYRGM3IAAMKKGSSDRYPQGSVNEAMKLPV
- 30 Sbjct: 361 NAVMLGSMFAGIDEAPGETEIQQRKPKTYRGM3IAAMKKGSSDRYPQGSVNEAMKLPV 420
- Query: 422 EGTGRVAYKGVADIVFQMLGIRSGMGTVGAANIKELDNAQFVEMSGAGLIESHPHD 481
EGTGRVAYKG+ +DIVFQMLGIRSGMGTVGA +I+ELG+NAQFVEMSGAGLIESHPHD
- 35 Sbjct: 421 EGTGRVAYKGAASDIVFQMLGIRSGMGTVGAADIQELHENAQFVEMSGAGLIESHPHD 480
- Query: 482 VQITNEAPNYSVH 494
VQITNEAPNYSVH
- Sbjct: 481 VQITNEAPNYSVH 493
- 35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1020

A DNA sequence (GBSx1089) was identified in *S. agalactiae* <SEQ ID 3139> which encodes the amino acid sequence <SEQ ID 3140>. This protein is predicted to be MutR. Analysis of this protein sequence

40 reveals the following:

Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.1841 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 50 >GP:AAD04237 GB:AF007761 MutR [Streptococcus mutans]
Identities = 51/215 (23%), Positives = 102/215 (46%), Gaps = 9/215 (4%)
- Query: 5 GKILKEAREDEKGISLSLAKSAQLSKSTLSRFENGETOIGIKDFIKALQTLLEVGVITINEV 64
G++ KEAR +G+ L +A+ LS S LS+FRNG+T+ DK I A+Q + +T+E
- 55 Sbjct: 9 GELYKELRMARGLIKDKIDARD-NLSVSQLSKFENGQTMIAADKLILAIQGIH--MTFSEP 65
- Query: 65 SILDSKVKAGISNTYDLEQJTLLESYKRNEDIMRIPSQQKQSCRIEENSVLKILAKLPIS 124
S ++ + ++L L++ +D ++ +I ++ + K++ K +
- 60 Sbjct: 66 SYAFTQYQESDLFKTKKGLVLTQTKDKIGLKKILKDYPTDTYTVNVRNLKLVIAEAVY 125
- Query: 125 NLGLENMRLPQDEINLVVTVINGVQYNDYFVKVICYFQDILPED--VILNKI-----SRMT 178

-1133-

+L + + +E + +YL + ++ + + IL +D V L K +
 Sbjct: 126 SLDSSEITNEKEPILTSYLAIEGWYELVLPNTLFLISDDLVFLGKAFVREDKLY 185
 Query: 179 KBQLPVSKSLVNLILKQVITALEKSDVKALVFAD 213
 +E + K +LI + +I +E S A F +
 Sbjct: 186 RLSSEKGRALVNLILVLSHHSFYHAQYFIE 220

There is also homology to SEQ ID 628.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1021

A DNA sequence (GBSx1090) was identified in *S. agalactiae* <SEQ ID 3141> which encodes the amino acid sequence <SEQ ID 3142>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -10.77 Transmembrane 269 - 285 (265 - 287)
 INTEGRAL Likelihood = -6.90 Transmembrane 33 - 49 (31 - 51)
 INTEGRAL Likelihood = -6.79 Transmembrane 182 - 198 (176 - 200)
 INTEGRAL Likelihood = -6.37 Transmembrane 117 - 133 (113 - 135)
 INTEGRAL Likelihood = -5.57 Transmembrane 240 - 256 (232 - 259)
 INTEGRAL Likelihood = -3.40 Transmembrane 223 - 239 (220 - 239)
 INTEGRAL Likelihood = -0.95 Transmembrane 56 - 72 (55 - 72)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5310 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3143> which encodes the amino acid sequence <SEQ ID 3144>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -10.99 Transmembrane 269 - 285 (264 - 286)
 INTEGRAL Likelihood = -8.76 Transmembrane 117 - 133 (112 - 135)
 INTEGRAL Likelihood = -7.70 Transmembrane 179 - 195 (174 - 200)
 INTEGRAL Likelihood = -4.83 Transmembrane 34 - 50 (32 - 52)
 INTEGRAL Likelihood = -4.46 Transmembrane 213 - 229 (211 - 230)
 INTEGRAL Likelihood = -4.14 Transmembrane 240 - 256 (232 - 259)
 INTEGRAL Likelihood = -0.59 Transmembrane 91 - 107 (91 - 108)
 INTEGRAL Likelihood = -0.32 Transmembrane 4 - 20 (4 - 20)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5394 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9181> which encodes the amino acid sequence <SEQ ID 9182>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -10.99 Transmembrane 259 - 275 (254 - 276)
 INTEGRAL Likelihood = -8.76 Transmembrane 107 - 123 (102 - 125)
 INTEGRAL Likelihood = -7.70 Transmembrane 169 - 185 (164 - 190)
 INTEGRAL Likelihood = -4.83 Transmembrane 24 - 40 (22 - 42)
 INTEGRAL Likelihood = -4.46 Transmembrane 203 - 219 (201 - 220)
 INTEGRAL Likelihood = -4.14 Transmembrane 230 - 246 (222 - 249)
 INTEGRAL Likelihood = -0.69 Transmembrane 81 - 97 (81 - 98)

-1134-

----- Final Results -----

bacterial membrane --- Certainty=0.539(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

An alignment of the GAS and GBS proteins is shown below.

Identities = 200/287 (69%), Positives = 244/287 (84%)

Query: 1 MEGLLIALIPMFANGSIGFVSNKIGGRPMQCTFCMTLQALLPATIIVLWPKQPEMTASLW 60
 +EG+ ALIPMF WGSIGFVSNKIGG+P+QCT GMT QALL++ VML +PEMT LW+
 Sbjet: 1 LEGIFYALIPMTWGSIGFVSNKIGGKPSQQTLMPTGALLSFLAVMLVLRPEMTLQML 60

Query: 61 FGILGGILSVGQNGQPFQAMKMGVSVMPLSSGQALVQGSILVGLVPHIEWTKPIQPILG 120
 FGILGG +WS+GQ QGF AM+VMGVSVANPLSSG+QLV GEL+G LVPHIEWT+P+QF++G
 Sbjet: 61 FGILGGFVNSIGQTCGFHAMQMGVSVMPLSSGQSLVLSLIGLVPHIEWTRPMQFVVG 120

Query: 121 LTALITLVIGFYFSSKRDVSEQLATHQEFSGKFATIAYSTGVLYSYAVLPHINIMKFDPM 180
 AL LL++GFYFSSK+D + + FSGKF + YST+GY+ YAVLPHINIMK+ +
 Sbjet: 121 SLALLLVIGFYFSSKQDDNAQVNHLMFSGKFRALTYSTIGVWIAVLPHINIMKFEVL 180

Query: 181 AVILPMAVGMCLGALCFMKFRVNFPAVVVQIMITGLAWGVGNVPMLLAAKAGLAIAFSP 240
 +VILPMAVGM LGAI FM F+++ + V+XN + GL+WG+GN+FMLLAA+KAGLAIAFSP
 Sbjet: 181 SVILPMAVGMVLGAITTFMSFKISIDQYVIRNSVGLLWIGINIFMLAASKAGLAIAFSP 240

Query: 241 SQLGVIITISGILFLGETTKTKKQKVVWGLICFVWGAILLGVKS 287
 SQLG IIST+GGILFLGETTKTKK +WVV GI+CF++GAILLG+VKS
 Sbjet: 241 SQLGAIISVGGILFLGETTKTKKEMRVVTGIICFIVGAILLGVVKS 287

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1022

A DNA sequence (GBSx1092) was identified in *S. agalactiae* <SEQ ID 3145> which encodes the amino acid sequence <SEQ ID 3146>. This protein is predicted to be recF protein (recF). Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2653(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3147> which encodes the amino acid sequence <SEQ ID 3148>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1677(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50

An alignment of the GAS and GBS proteins is shown below.

Identities = 248/364 (68%), Positives = 300/364 (82%), Gaps = 1/364 (0%)

Query: 1 MWIKNISLKHRYNYEERQVDFSPMLNIFIGRNAQGGKTNFLEAIYFLALTRSHRTRSDKEL 60
 MWIK + LKHRYNY+ FS LN+FIG NAQGGKTNFLEAIYFL+LTRSHRTR+DKEL
 Sbjet: 1 MWIKELELKHRYNYDHLLASPSGGLNVFTGNNAQGGKTNFLEAIYFLSLTRSHRTRADKEL 60

55

-1135-

Query: 61 VHPKHVDVQITGEVIRKSGHNLIDQLSEKGRITKVNLHQAQLSDYIGAMTVLFAPE 120
+HP H V +TG++ R SG ++L+I LS+KGR+TK+N LKQAQLSDYIG M VVLFAPED
Sbjct: 61 IHPDHSVSLTGKIQRIISGTVLEINLSKGRVTKINALQAQLSDYIGMTVMVLFAPED 120

Query: 121 LQLVKGAPSLRRKFLDIDIQIKPTYLARLSNNHVLKQRTYLYKITTNNVDKTFVLVLE 180
LQLVKGAPSLRRKF+DID+QIKP YL+ELG+YNNHVLKQRM+YK+ +D FL VLE
Sbjct: 121 LQLVKGAPSLRRKFIDIDLQIKPVYLSKLSHNVHVLKQRM+YKSAQQIDAAFLVLE 180

Query: 181 QLADYGSRVIEHRFDIQLANDEADKHYYIISTELHLSIHYKSSIEFTDKSSIRSHPLN 240
QLA YG+RV+KHR DFI AL EA+ HH IS IE LS+ Y+SS+ F K++I + FL+
Sbjct: 181 QLASYGARVMIEHRIDFINALEKANTHQAISNGLESLSVQSSVVFDDKKNIVQQFLH 240

Query: 241 QLSKSHSRDIFKNTSIGPHRIDITFFINDINATFASOQOQSLLISLKLAEIELIKVT 300
QL K+H +D F+KNTS+GPHRD++ P+IN +NA FASOQ RSLLSLKL+AE+ L+K +T
Sbjct: 241 QLEKSHQKDFPRNTSVGPHRELAFFYINGNNANFASOQOQSRSLLSLKLAEVSLMGLT 300

Query: 301 NDYPIILLDDVMSELDNHRQLKLEGG-IKENVQTFITTTSLHLSALPDQLKIFNVSDGT 359
D PILLDDVMSELDN RQ KLE IKENVQTFITTTSL+HLS LP+ ++IP+V+ GT
Sbjct: 301 GDNPIILLDDVMSELDNTRQTKLETVIKENVQTFITTTSLHLSALPDQLKIFNVKTGT 360

Query: 360 ISIN 363
+ I+
Sbjct: 361 VOID 364

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1023

A DNA sequence (GBSx1093) was identified in *S. agalactiae* <SEQ ID 3149> which encodes the amino acid sequence <SEQ ID 3150>. Analysis of this protein sequence reveals the following:

Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1807 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA61548 GB:U89367 orf121 [Lactococcus lactis]
Identities = 56/116 (48%), Positives = 74/116 (63%), Gaps = 3/116 (2%)

Query: 3 YKLFDEYITLQSLKKRIGITQSGGAIKKFLADR--VLPNGDLNREGKKLRGLDITIP 60
Y LP+EVITL LKKE+G+I +GG K FLA+N + +NG+ ENREGKKLR GD++ P
Sbjct: 4 YILFDEYITLQSLKKRIGITQSGGAIKKFLADR--VLPNGDLNREGKKLRGLDITIP 60

Query: 61 DONIKIIRKPSDQRIERNIARQKRVSAIVKMNKNTNGKSKTSKKVVRFPFG 116
++++ + I+E E AE+ RV AIVK+MN NK K P RFPFG
Sbjct: 64 TFDLKVTFQADADAKHEIRAEKAREARVKAIVKMNAE-NKTTKPAKKAPRRFPFG 118

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3151> which encodes the amino acid sequence <SEQ ID 3152>. Analysis of this protein sequence reveals the following:

possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0493 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-1136-

An alignment of the GAS and GBS proteins is shown below.

Identities = 74/136 (54%), Positives = 94/136 (68%), Gaps = 20/136 (14%)

```

5   Query: 1 MDYKLPDEYITLQSLIKETIGIQSGGAIKKFLADNRVLPNGDLENRRGKKLRGLDITIP 60
      M YKLF E+ITLQ+LHKE+GIQSGGAIK FLA+ VLPNG+ E RRGKK+R+GD I++P
      Sbjct: 9 MIYKLPTEFITLQALLKELGIQSGGAIKGFLAETTVLPNGEDEKRRGKKIRVGDKISLP 68

      Query: 61 DQNIHIIIRKPSDQIEERNEIARQKRVSAIVKQ+MKYTNKGRSK-----TSKK---- 110
      DQ+ I I I +PS +E E+ E+ARK RV+A+VK+M+ K SK T+KK
10   Sbjct: 69 DQDLIITIVEPSQREKEQFAREMAKTRVAALVKQ+M+QANKKTSKSHNNQST+TKKSLRA 128

      Query: 111 -----PVRFPFG 116
      PVRFPFG
15   Sbjct: 129 TKKTGKKTAPVRFPFG 144

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1024

A DNA sequence (GBSx1094) was identified in *S. agalactiae* <SEQ ID 3153> which encodes the amino acid sequence <SEQ ID 3154>. Analysis of this protein sequence reveals the following:

```

Possible site: 47
>>> Seems to have no N-terminal signal sequence
      INTEGRAL Likelihood = -1.86 Transmembrane 269 - 285 ( 267 - 285)

```

```

25   ----- Final Results -----
      bacterial membrane --- Certainty=0.1744(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3155> which encodes the amino acid sequence <SEQ ID 3156>. Analysis of this protein sequence reveals the following:

```

Possible site: 48
>>> Seems to have no N-terminal signal sequence

```

```

35   ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3008(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 227/413 (54%), Positives = 309/413 (73%)

```

Query: 1 MKIVGVSLHLIKNQQFKTNHLTPFRSGDFNNKTVARSLVACMLVTANAKYPKVQVPRE 60
      MKIV+GV LHLIK +QFKTNH+TPFRSGD N KTVAA+ LVACML TAN YP V++PRE
45   Sbjct: 1 MKIVQGVQLHLIKTKQFKTNHLTPFRSGDLNQKTVAKKVLVACMLATANECPTVCPVPRE 60

      Query: 61 KLASLNGASLSTKSTKGLVHVIDIVFVNKPTLEQRNIVQIITPLEMDLSPSLISL 120
      KLA LKASLST + TKGLVHVIDIDI F++ + E I++++I FL+D+LSPSL+SL+
50   Sbjct: 61 KLARLNGASLSTNVLTKGLVHVIDIDITPIQDRYACNGEKILDENIQPLDKLSPSLIST 120

      Query: 121 EOYQTSIPDTEKKNLIQVLEADIEDNPFYSSDLAKSLFYNNKTLRLPKYGTASLIVESNS 180
      QYQ +F+TEK NLI Y+E+D ED+FY S L +K LFY NK L++ +YG+ L+ E+
      Sbjct: 121 AQYQKVFETEKKNLNIYTSREDSPFYSSLKVKELFYCNKMLQMSYEGSPSLIAKSTA 180

55   Query: 181 PTAYGEFQMKLKEQDQIDFVCGDDPYRMICAFNRMAFEPRIKVLAFDYTYQTYENITRSQ 240
      +T+YQEF KML EDQ+DIF++GDDPYR++Q ++ + R+K L F + Q NI +
      Sbjct: 181 YTSYQRFPHMLNEDQIDIFILGDDPYRVVQLIHQFPLDNKNENLNPHLQNSVNIKES 240

      Query: 241 VEDKIDVNSIMQLAYHLPTTYKEDDYFALIVENGILGAFASLSLPTREKQGLATTIGS 300

```

-1137-

```

+ E + V + QSI + QLAY P + DY + AL ++ NGL G + FAHS LF + IRE ++ GLAY + IG
Sbjct: 241 IIRKRAVHQSILQIAYEFVFPVQGRDYVAIVLLNGLGSPFAHSRLFIKIRREGLAYSIGC 300

Query: 301 QFDSFTGLFTIYAGIDKNERFLKLINKQFNNTKMGFRFSSTLLKQTKDILDMNYVLASD 360
+ FDS + TGLF IY GID ++ R + L + LI ++ N IKMGFRFS L + K + T + + L N + L + D
Sbjct: 301 RFDSTYGLFETIYTGIDSQHRITLQLIIQELNALIKMGFRFSEQLIKKTSMLNNALLSED 360

Query: 361 NPKVIVDHITYEHYLDQPHTSALFIDKVDVTKSDIVSVATKLLKQAFYFLEG 413
K I ++ IY Y + D ++ + I V ++ V K + DI + VA LKLG YFLEG
Sbjct: 361 YNKNILIERIYRASYDSSYSIKNWKGVNEVKNKADILKVANLLKLTQVYFLEG 413

```

SEQ ID 3154 (GBS400) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 76 (lane 2; MW 49.2kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 84 (lane 3; MW 74kDa) and in Figure 177 (lane 6; MW 74kDa).

GBS400-GST was purified as shown in Figure 217, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1025

A DNA sequence (GBSx1095) was identified in *S. agalactiae* <SEQ ID 3157> which encodes the amino acid sequence <SEQ ID 3158>. Analysis of this protein sequence reveals the following:

```

Possible site: 50
>>> Seems to have no N-terminal signal sequence

```

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3473 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3159> which encodes the amino acid sequence <SEQ ID 3160>. Analysis of this protein sequence reveals the following:

```

Possible site: 45
>>> Seems to have no N-terminal signal sequence

```

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4298 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 207/424 (48%), Positives = 276/424 (64%), Gaps = 3/424 (0%)

Query: 5 KITYNQLQHEVVKLTLESGLANVLIPKPSFKTGVGLVITANPGSLHTKYTRNGCVHEYIPAG 64
KI Y N+ E++Y + LE+GL VY I K F E +LT FQSL K T + PAG
Sbjct: 6 KINYPNIDEDLYYVKLENGILVYVTFIKKIGFLEKRMVLTVGPGSLONKLTVDDESRLDAPAG 65

Query: 65 IAHFLHKLPLFELDKGQDAATQFTKYGASNAFTTVEKTSFFPSTISHITNCLDILLDFVL 124
IAHFLHKLPLF + G D + +ET+ GAR+NAFTTF++TSP+PST S L++L FVL
Sbjct: 66 IAHFLHKLPLFDESQGDISLQFTQLGAEINAFITTNKTSFFPSTASKPQSNLELQYFVL 125

Query: 125 TTNFTRESITTEKEDIKIKREIKYQDPPEYRKYGVGLSNLYPNSPLAFDIAGDYQSIQIT 184
+ N T+ES++EK II QEI+MYQDD +YR Y G+L NL+P + LA DIAG SI +IT
Sbjct: 126 SANITDESVSREKKIIGQEDIMYQDDADYRAYBSGLIQLNLFKPTSLANDIGSKASIQKIT 185

Query: 185 LTDLQENHKDFYOLS+SN+NLVNGQFSPQLIITYLQNSHFTSY--SQNIDRDSISLEPVI 242
L+ +H FYQ +NM+L +VG K +Q+ SY ++ D + PVI

```

-1138-

Sbjct: 186 KILLSTHTYFYQPTNMSLFTVGLDIDETFLAIQRFPQTLYSPDRKRVTVDPHLYPVI 245
 Query: 243 KNNSCMTVTKPKLAIGYRKSNHMHGSLYKKEKIGLQFPFAMLGWTSTINGWYSSQI 302
 K++S M VT KL +G+R + S L +I L+L+P +ML+GWIS I YR G+I
 Sbjct: 246 KSSSDVMDVITAKLVVGGFVGLTLTQHSLLTYRIAKLFLSMLIGWTSKIYHTLYEDGKI 305
 Query: 303 DDSFDIEIEVHPDFECVLIISLDTTEPIAFSTQIRLLKNAQSSDLTSHLNVKRELYG 362
 DDSFD++R+H +F+ V+ISLDT EPIA S +R L S +T EL +K+R+YG
 Sbjct: 306 DDSFDVDEIHHNFQFVLIISLDTPEPIAMSNYIRQKATIKISKEPTNEHLMLKKEVYG 365
 Query: 363 DFLRSLDSIENLAMOQFVTVLYDG-KTMYLDLPSIVERDLLEDVITIGKDFLONADTSDFV 421
 DF++SLDSIE+L QF YL D K Y D+P I+E L L+DV+TIGK F + AD SDF
 Sbjct: 366 DFIQSLDSIEHLTHQFSLYLSDSKRTYFDIPKIIIRLITLKDVTIGKAFFEKADASDFT 425
 Query: 422 IFPK 425
 +FPK
 Sbjct: 426 VFPK 429

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1026

A DNA sequence (GBSx1096) was identified in *S. agalactiae* <SEQ ID 3161> which encodes the amino acid sequence <SEQ ID 3162>. This protein is predicted to be phosphatidylglycerophosphate synthase (pgsA). Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.17	Transmembrane	17 - 33 (14 - 39)
INTEGRAL	Likelihood = -3.77	Transmembrane	92 - 108 (88 - 108)
INTEGRAL	Likelihood = -2.87	Transmembrane	144 - 160 (142 - 162)
INTEGRAL	Likelihood = -1.65	Transmembrane	42 - 58 (42 - 59)

 ----- Final Results -----
 bacterial membrane --- Certainty=0.4270 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10293> which encodes amino acid sequence <SEQ ID 10294> was also identified.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3163> which encodes the amino acid sequence <SEQ ID 3164>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -6.64	Transmembrane	76 - 92 (72 - 102)
INTEGRAL	Likelihood = -5.36	Transmembrane	136 - 152 (131 - 164)
INTEGRAL	Likelihood = -2.34	Transmembrane	98 - 114 (97 - 114)

 ----- Final Results -----
 bacterial membrane --- Certainty=0.3654 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 145/180 (80%), Positives = 160/180 (88%)
 Query: 8 MTKKENIENLLTVRRIIMLPFLVLTSTVTSTTHVIAVIAFPAIASLTDLGDLARKW 67
 M+KKKENIENLLT+VRI MIP P+ +TS + WHI AA++FAIAS TTYLDGDLARKW
 Sbjct: 1 MTKKENIENLLTVRRIAMLPFLTSSNKVGVHIAVIAFPAIASPTDLYLDGDLARKW 60

-1139-

Query: 68 VVTNFGKPADPLADKMLVMSAFIMLVGLDLAPANVSATIIICRELAVTGLRLLLVETGTV 127
 V +NFGKPADPLADKMLVMSAFIMLVGL L PANVSA+IICRELAVTGLRLLLVETGG V
 Sbjct: 61 VASNFGKPADPLADKMLVMSAFIMLVGLGLVPAVNSAVIICRELAVTGLRLLLVETGSKV 120

5 Query: 128 LAAAMPGKIKTATQMFVIFLLVHMVTIGNIMLYALFPTIYSGDYFKGAGFLFKDTFK 187
 LAAAMPGKIKTATQM ++I LL HW+ LGN+LYTALFPTIYSGDYFKGA FLFKDTFK
 Sbjct: 121 LAAAMPGKIKTATQMLSIILLCHWIFLGNVLYALFPTIYSGDYFKGASFLFKDTFK 180

A related GBS gene <SEQ ID 8705> and protein <SEQ ID 8706> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 4
 SRCFLG: 0
 MCg: Length of UR: 9
 Peak Value of UR: 3.03
 Net Charge of CR: 1
 MCg: Discrim Score: 6.36
 GVH: Signal Score (-7.5): -0.400001
 Possible site: 48
 >>> Seems to have a cleavable N-term signal seq.
 Amino Acid Composition: calculated from 49
 ALOM program count: 2 value: -3.77 threshold: 0.0
 INTEGRAL Likelihood = -3.77 Transmembrane 85 - 101 (81 - 101)
 INTEGRAL Likelihood = -2.87 Transmembrane 137 - 153 (135 - 155)
 PERIPHERAL Likelihood = 1.27 109
 modified ALOM score: 1.25
 icml HYPID: 7 CFP: 0.251

*** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.2508 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1027

A DNA sequence (GBSx1097) was identified in *S. agalactiae* <SEQ ID 3165> which encodes the amino acid sequence <SEQ ID 3166>. This protein is predicted to be ABC transporter ATP-binding protein (potA). Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1805 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP: AAC61484 GB: AF082738 ABC transporter ATP-binding protein
 [Streptococcus pyogenes]
 Identities = 201/279 (72%), Positives = 231/279 (82%)

Query: 1 MTNIIITVNNLFPRYDSNQTHYQLSNVSPHVKGQEWLSIIGHNSGKSTTVRLIDGLLEPE 60
 M+ II + + F Y +Q L+ VSPHVKGQEWLSIIGHNSGKSTT+RLIDGLLE E
 Sbjct: 18 MSATIELKKVTPYHIDQEKFTLDGVSFHVKGQEWLSIIGHNSGKSTTVRLIDGLLEPE 77

Query: 61 SQGIIDIGQLTFEDNVWELHKKIQVNPQNPQNGVATVEDDVARGLENGEIPFKDKMER 120
 SG II+DG LT NVWB+RHKIGNVFPQNPQNGVATVEDDVARGLENGE I+D+KER

-1140-

- 5
10
15
- Subjct: 78 SGSIIVDGDLLTITNVWEIRKHIGMVQNPQNFVGAIVEDDVAFGLENKGIHEDIKER 137
- Query: 121 VDQALDLVGMSEFKMREPARLSGGQKQKQVVAIGAVNMRPOVILLDEATSMLDPGRLLEL 180
V+ AL+LVGM FK +EPARLSGGQKQKQVVAIGAVNMRPOVILLDEATSMLDPGRLLEL 180
- Subjct: 138 VNHAEELVGMQMFKEKEPARLSGGQKQKQVVAIGAVNMRPKIILLDEATSMLDPGRLLEL 197
- Query: 181 RTIRAIRQKMYLTVISITHDLDEVALSDRVIVMKNKVESTSTPKALPGRGNRLISGLD 240
+TI+ IR Y LTVISITHDLDEVALSDRV+VMK+G+VESTSTP+ LF RG+ L+ LGLD 240
- Subjct: 198 KTKIKNIRDYQLTVISITHDLDEVALSDRVLMKDGQVESTSTP+EQLFARGDELLGLD 257
- Query: 241 VPPTSRMLAELAANGLDIGTEYLTEKLEBQWLNLNLM 279
+PPT+ ++ L G I YLTKLE QL +L KM 279
- Subjct: 258 IPFTTSVVQMLQEGGYPIDYGYLTKELNQLQLISLM 296
- A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3167> which encodes the amino acid sequence <SEQ ID 3168>. Analysis of this protein sequence reveals the following:
- Possible site: 19
>>> Seems to have no N-terminal signal sequence
- Final Results -----
bacterial cytoplasm --- Certainty=0.2235 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
- REG motif: 247-249

An alignment of the GAS and GBS proteins is shown below.

- Identities = 200/279 (71%), Positives = 231/279 (82%)
- 30
35
40
45
50
- Query: 1 MTNIIITVNNLFFPKYDSNCTHYQLNVSFHVKGSEWLSIIHGNSSGKSTTVRLIDGLLEB 60
M+ II + + F Y +Q L+ VSFHVKGSEWLSIIHGNSSGKSTTVRLIDGLLEB 60
- Subjct: 18 MSAILLEKKTFTNYHKDQKPTLDGVSFHVKGSEWLSIIHGNSSGKSTTVRLIDGLLEB 77
- Query: 61 SGOIILDQQLTENVWELAKHIGMVQNPQNFVGAIVEDDVAFGLENKGIHEDIKER 120
SG II+DG LT NVWE+RHKIGMVQNPQNFVGAIVEDDVAFGLENKGI +D+KER 120
- Subjct: 78 SGSIIVDGDLLTITNVWEIRKHIGMVQNPQNFVGAIVEDDVAFGLENKGIHEDIKER 137
- Query: 121 VDQALDLVGMSEFKMREPARLSGGQKQKQVVAIGAVNMRPOVILLDEATSMLDPGRLLEL 180
V+ AL+LVGM FK +EPARLSGGQKQKQVVAIGAVNMRPOVILLDEATSMLDPGRLLEL 180
- Subjct: 138 VNHAEELVGMQMFKEKEPARLSGGQKQKQVVAIGAVNMRPKIILLDEATSMLDPGRLLEL 197
- Query: 181 RTIRAIRQKMYLTVISITHDLDEVALSDRVIVMKNKVESTSTPKALPGRGNRLISGLD 240
+TI+ IR Y LTVISITHDLDEVALSDRV+VMK+G+VESTSTP+ LF RG+ L+ LGLD 240
- Subjct: 198 KTKIKNIRDYQLTVISITHDLDEVALSDRVLMKDGQVESTSTP+EQLFARGDELLGLD 257
- Query: 241 VPPTSRMLAELAANGLDIGTEYLTEKLEBQWLNLNLM 279
+PPT+ ++ L G + YLTKLE QL +L KM 279
- Subjct: 258 IPFTTSVVQMLQEGGYPIDYGYLTKELNQLQLISLM 296
- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1028

A DNA sequence (GBSx1098) was identified in *S.agalactiae* <SEQ ID 3169> which encodes the amino acid sequence <SEQ ID 3170>. Analysis of this protein sequence reveals the following:

- 55
60
- Possible site: 49
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.27 Transmembrane 154 - 170 (154 - 170)
- Final Results -----
bacterial membrane --- Certainty=0.1107 (Affirmative) < succ>

-1141-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:CAB11922 GB:Z99104 similar to ABC transporter (ATP-binding
 protein) [Bacillus subtilis]
 Identities = 141/242 (58%), Positives = 188/242 (77%), Gaps = 1/242 (0%)

10 Query: 16 TPFEGRALFDVNLKIEDASVYAFIHTGSGKSTIMQLHGLHIPTKGEVIVDDPSIKAGD 75
 TPFE AL+D+N I++ SY A IGHGTGSGKST++Q LKGL PTEG++ + I+AG
 Sbjct: 3 TPFERIALYDINASIKESYVAVIGHTGSGKSTLQHANGLLKPTGQISLGSVVIQAGK 62

15 Query: 76 KNKEIKFIRQKVLGVFPFESQLFETVYLKQVARGPQNFQISQIBARLIAREKLRVIGIS 135
 KNK+K +R+KVG+VFQFPE QLFETVYLDK++RGP NFG+ + +RE+ A E L+LVG+S
 Sbjct: 63 KNKDLKLRKKKVGIVFPQFPEHQLFETVYLKDISPGPNFVGKKEDARQKAREMLQLVGLS 122

20 Query: 136 EDLFDKNPFELSGGQMRRVAIAGIAMEPKVLVLDEPTAGLDPKRKELMTLFPKHLHKG 195
 E+L D++PFELSGGQMRRVAIAG+LAM+P+VLVLDEPTAGLD+GRKE+M +F LH+G
 Sbjct: 123 EELDRSPFELSGGQMRRVAIAGVLANDEPVVLVLEPTAGLDPRGRKEIMDMFYELHGR 182

25 Query: 196 -MTIVLVTHLMDDVADYADYVYVLKAGVILSGQPKQIFQBEVLSESKQLGVPKTIKPAQ 254
 +T +LVTH M+D A YAD + V+ G + SG P+ +F + E + L +P+ K+ +
 Sbjct: 183 NLTTILVTHSMEDAAAYADEMIVHKGITIQASGSPDLFLKGEEMAGWGLDLPETIKFQR 242

25 Query: 255 RL 256
 L
 Sbjct: 243 HL 244

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3171> which encodes the amino acid
 30 sequence <SEQ ID 3172>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> There to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.27 Transmembrane 154 - 170 (154 - 170)

35 ----- Final Results -----

bacterial membrane --- Certainty=0.1107 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

40

The protein has homology with the following sequences in the databases:

>GP:CAB11922 GB:Z99104 similar to ABC transporter (ATP-binding
 protein) [Bacillus subtilis]
 Identities = 146/259 (56%), Positives = 187/259 (71%), Gaps = 2/259 (0%)

45 Query: 16 TPFEGRALFNNILDIIDGVSYAFIHTGSGKSTIMQLHGLNVPPTIIVSDKQDITHES 75
 TPFE AL++LN I +GSY A IGHGTGSGKST++Q LKGL PT G +S+ I
 Sbjct: 3 TPFERIALYDINASIKESYVAVIGHTGSGKSTLQHANGLLKPTGQISLGSVVIQAGK 62

50 Query: 76 KNKEIKSIRKHVQLGVFPFESQLFETVYLKQVARGPQNFQVSPSEAEALAREKLRVIGIS 135
 KNK+K +RK VG+VQFPE QLFETVYLDK++RGP NRGV R+AR ARE L I+VG+S
 Sbjct: 63 KNKDLKLRKKKVGIVFPQFPEHQLFETVYLDKDISPGPNFVGKKEDARQKAREMLQLVGLS 122

55 Query: 136 ENLFEKNPFELSGGQMRRVAIAGIAMEPKVLVLDEPTAGLDPKRKELMTIIPKHLHKG 195
 E L +++PFELSGGQMRRVAIAG+LAM+P+VLVLDEPTAGLD+GRKE+M +F +LH+G
 Sbjct: 123 EELDRSPFELSGGQMRRVAIAGVLANDEPVVLVLEPTAGLDPRGRKEIMDMFYELHGR 182

60 Query: 196 -MTIVLVTHLMDDVADYADYVYVLKAGVILSGQPKTIQQQVSLLEKQLGVPKTIKPAQ 254
 +T +LVTH M+D A YAD + V+ KG I SG P+ +F + + L +P+ K+ +
 Sbjct: 183 NLTTILVTHSMEDAAAYADEMIVHKGITIQASGSPDLFLKGEEMAGWGLDLPETIKFQR 242

Query: 255 RL-VDRGIPISSLPTLHE 272
 L G+ + +T+R+
 Sbjct: 243 HLEALGVKRFNPEMLTIED 261

An alignment of the GAS and GBS proteins is shown below.

Identities = 218/280 (77%), Positives = 241/280 (85%)

```

5  Query: 1  NGIEFKNVSYTYQAGTFPFEGRALFDVNIKIDASYTAFTIGHTGSGKSTIMQLLGLHPT 60
      N I +NVSYTYQAGTFPFEGRALF+NL I D SYTAFTIGHTGSGKSTIMQLLGLH+PT
      Sbjct: 1  NSINLQNVSYTYQAGTFPFEGRALFNINLIDGSSYTAFTIGHTGSGKSTIMQLLGLHVPT 60

10  Query: 61  KGEVTVDDFSIKADIKKIKIPKIRQKVLVQFPFESQLPRTVLKDVARGPQNGISQIE 120
      G V V D I KKEIK IR+ VGLVQFPFESQLPRTVLKDVARGPQNG+G S E
      Sbjct: 61  TGIVSVDKQDITNHSIKKIKIKIRKHVGLVQFPFESQLPRTVLKDVARGPQNGVSPES 120

15  Query: 121  AERLAEEKLIRVIGISEDLPKNPFELSGGQMRVAIAGILAMEPKVLVLDEPTAGLDPEK 180
      AE LA EKL IGVISE+L+KNEPFELSGGQMRVAIAGILAM+PKVLVLDEPTAGLDPEK
      Sbjct: 121  AEALAREKLALGVIGISELNFKNPFELSGGQMRVAIAGILAMQKVLVLDELPTAGLDPEK 180

20  Query: 181  RKELMTLFLPNLHKGMTIVLVTHLMDVDVADYVYVLEAGKVTLSGQPKQIPQEVLE 240
      RKELMT+PK LH+ GMTIVLVTHLMDVDA+YAD+VYVLA GK+ LSG+PK IPQ+V LLE
      Sbjct: 181  RKELMTIFKKLHGSGMTIVLVTHLMDVDVANYADFYVLDKGIILSGKPTIPQVSLLE 240

      Query: 241  SKQLGVPKTKFAQRILSHKGLNLPSLPITINEFVAIKHG 280
      KQLGVPK+TK AQLR +G+ +SLPIT+ E E +KHG
      Sbjct: 241  KKQLGVPKVTKLARLVDRGIPISLSPLITLELRVLKHG 280

```

25 SEQ ID 3170 (GBS401) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 76 (lane 3; MW 34.4kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 84 (lane 4; MW 59kDa).

GBS401-GST was purified as shown in Figure 218, lane 2.

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1029

A DNA sequence (GBSx1099) was identified in *S. agalactiae* <SEQ ID 3173> which encodes the amino acid sequence <SEQ ID 3174>. Analysis of this protein sequence reveals the following:

```

35  Possible site: 43
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL Likelihood = -10.46 Transmembrane 47 - 63 ( 25 - 69)
      INTEGRAL Likelihood = -8.81 Transmembrane 252 - 268 ( 249 - 269)
      INTEGRAL Likelihood = -7.91 Transmembrane 116 - 132 ( 110 - 141)
40  INTEGRAL Likelihood = -4.25 Transmembrane 29 - 45 ( 25 - 46)
      INTEGRAL Likelihood = -2.55 Transmembrane 77 - 93 ( 77 - 95)
      INTEGRAL Likelihood = -0.43 Transmembrane 199 - 215 ( 199 - 215)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.5182(Affirmative) < succ>
45  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8707> which encodes amino acid sequence <SEQ ID 8708> was also identified. Analysis of this protein sequence reveals the following:

```

50  Lipop: Possible site: -1 Crend: 7
      SRCFLG: 0
      McG: Length of UR: 8
      Peak Value of UR: 0.65
      Net Charge of CR: 1
55  McG: Discrim Score: -10.55
      GvH: Signal Score (-7.5): 1.45

```


-1143-

Possible site: 37
 >>> Seems to have no N-terminal signal sequence
 Amino Acid Composition: calculated from 1
 ALIGN program count: 6 value: -10.46 threshold: 0.0

5	INTEGRAL	Likelihood = -10.46	Transmembrane	41 - 57 (19 - 63)
	INTEGRAL	Likelihood = -8.81	Transmembrane	246 - 262 (243 - 263)
	INTEGRAL	Likelihood = -7.91	Transmembrane	110 - 126 (104 - 135)
	INTEGRAL	Likelihood = -4.25	Transmembrane	23 - 39 (19 - 40)
	INTEGRAL	Likelihood = -2.55	Transmembrane	71 - 87 (71 - 89)
10	INTEGRAL	Likelihood = -0.43	Transmembrane	193 - 209 (193 - 209)
	PERIPHERAL	Likelihood = 0.79	90	

modified ALIGN score: 2.59
 icml HYPIID: 7 CFP: 0.518

15 *** Reasoning Step: 3

----- Final Results -----

	bacterial membrane ---	Certainty=0.5182 (Affirmative) < succ>
	bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>
20	bacterial cytoplasm ---	Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB11923 GB:299104 ybaF [Bacillus subtilis]
 Identities = 133/263 (50%), Positives = 191/263 (72%)

25 Query: 7 MDKLLGRYIPGNSLIHKLDPRSKLLAMLLFIIIVFKNHVVTVIVPIFTLVIVLSQI 66
 MD +I+G+Y+PG SL+H+LDP+KL+ + LF+ IVF ANNV T ++ +FT+ +V L+ ++
 Sbjct: 2 MDSMIIGKYVPGTSLVHKLDPRTKLITIFLFCVIFLANNVQTYALLGLFTIGVSLTRV 61

30 Query: 67 KFGYFNGIKFVGVIILPTTLFQALFAQQGVIFSWFIPSTLSGLQQAALINRFLVLI 126
 F6+ G+KP++ I+LFT L +L G +IF + GL Q I +RFV +I
 Sbjct: 62 PFSFLMKGLKPIIIVLPTFLHLIHLTHEQPIIPQIGFSRVYSGILVQGISIFSRFVYLI 121

35 Query: 127 FSTLLTTLTTTFLSLADAVESLLKPLEVLRVPAHEIGLMSLSLRFVPTIMDDTTRIMNA 186
 +TALTLTTT+ + D +E LL PL+ L+P HE+ LM+S+SLRF+FTLM++T +IM A
 Sbjct: 122 LITTLTTLTTTPIBITDMEQLNPLKKLKLPHKLALMMSISLRIFTIMBETDKIMKA 181

Query: 187 QRARGVDFGEGNLHKVKSIIPIILPLFASFPRADALAJAMEARGYQGGANRSKYRLK 246
 Q ARGVDF G + +VK+I+P+L+PLF S+FKRA+ LA+AMEARGYQGG R+KYR L
 40 Sbjct: 182 QMARGVDFGCPVSRKVAIVPLLVPLFVSAFKRAELAVAMEARGYQGBORTKYRKLIV 241

Query: 247 WTVRDTFSLILMLLGLSLPLK 269
 WT +DT I+ +++L LP L+
 Sbjct: 242 WTGKDTSVIVSLIVLAALLFSLR 264

45

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3175> which encodes the amino acid sequence <SEQ ID 3176>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

50	INTEGRAL	Likelihood = -9.50	Transmembrane	246 - 262 (243 - 265)
	INTEGRAL	Likelihood = -9.34	Transmembrane	110 - 126 (103 - 135)
	INTEGRAL	Likelihood = -6.69	Transmembrane	41 - 57 (40 - 58)
	INTEGRAL	Likelihood = -2.81	Transmembrane	23 - 39 (21 - 40)
	INTEGRAL	Likelihood = -1.01	Transmembrane	62 - 78 (62 - 78)
55	INTEGRAL	Likelihood = -0.27	Transmembrane	193 - 209 (193 - 209)

----- Final Results -----

	bacterial membrane ---	Certainty=0.4800 (Affirmative) < succ>
	bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>
60	bacterial cytoplasm ---	Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP: CAB11923 GB:299104 ybaF [Bacillus subtilis]
 Identities = 138/263 (52%), Positives = 195/263 (73%)

-1145-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3181> which encodes the amino acid sequence <SEQ ID 3182>. Analysis of this protein sequence reveals the following:

Possible site: 31

5 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

10 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 116/233 (49%), Positives = 140/233 (59%), Gaps = 39/233 (16%)

15 Query: 9 KLNVKGHHLAYGAILTVALFSCILAVMVIFKSSQVITSLSKADKRVVAKKSK----- 61
K N+K+ + +G LVAL ILH+ F S T+S +K + ++ K
Sbjct: 4 KENLKQRYFNFQ---LVALVALTILAILFAPSSKNADTKSYAKKSESKMVTIDKPIKNMHA 60

20 Query: 62 MTKATSKSKVEDVVKQAPKPSQASNRAPKSSSQSTEANSQQQVITASKEAAVEQAVVTNFT 121
+TK SK K + + P P+ ++ AP T +E V Q VT
Sbjct: 61 ITKEESKEKAKSLASEPIPTVENSAP-----TIVREVVVQQEV----- 101

25 Query: 122 ATSQAQQAAYAVTETTYTPAQHQTSQQVLSNGNTAGAGSAAAAQMAAATGVPGSTWEHII 181
Q V+ Y P + VLSNGNTAG +GS AAAQMAAATGVPGSTWEHII
Sbjct: 102 ----CTVQOVVSAYNP-----NNVLSNGNTAGIVGSQAAQMAAATGVPGSTWEHII 151

Query: 182 ARESNGNPNVANASGASGLFQTFPGWGSTATVQDVNSAIKAYRAQLSANGY 234
ARESNGNPN ANASGASGLFQTFPGWGSTATV+QDVN+A+KAY AQLSANGY
Sbjct: 152 ARESNGNPNANASGASGLFQTFPGWGSTATVEDQVNAALFAYSRAQLSANGY 204

A related GBS gene <SEQ ID 8713> and protein <SEQ ID 8714> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8

McQ: Discrim Score: 2.48

GVH: Signal Score (-7.5): -3.74

Possible site: 45

>>> Seems to have an uncleavable N-term signal seq

ALOM program count: 1 value: -12.05 threshold: 0.0

INTEGRAL Likelihood = -12.05 Transmembrane 22 - 38 (16 - 43)

PERIPHERAL Likelihood = 4.29 156

modified ALOM score: 2.91

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.5819 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

50 The protein has homology with the following sequences in the databases:

61.8/68.7% over 114aa

Staphylococcus aureus

GP|7959131| secretory protein SAI-B Insert characterized

55 ORF01057(664 - 1002 of 1302)

GP|7959131|dbj|EAA95959.1||AB042839(119 - 233 of 233) secretory protein SAI-B
{Staphylococcus aureus}

%Match = 15.1

%Identity = 61.7 %Similarity = 68.7

60 Matches = 71 Mismatches = 34 Conservative Sub.s = 8

438 468 498 528 558 588 618 648

-1146-

[illegible]

SEQ ID 3180 (GBS25) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 5; MW 25kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 11; MW 50kDa), Figure 63 (lane 6; MW 50.3kDa), Figure 66 (lane 6; MW 50kDa) and in Figure 175 (lane 8 & 9; MW 50kDa).

Purified GBS25-GST is shown in Figure 9A, Figure 193 (lane 11) and Figure 210 (lane 5).

The purified GBS25-GST fusion product was used to immunise mice (lane 1+2+3 products; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 95B), FACS (Figure 95C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS 25 bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1031

A DNA sequence (GBSx1103) was identified in *S. agalactiae* <SEQ ID 3183> which encodes the amino acid sequence <SEQ ID 3184>. This protein is predicted to be L-serine dehydratase 1 (sdaA-2). Analysis of this protein sequence reveals the following:

```

Possible site: 61
>>> Seems to have no N-terminal signal sequence
35  INTEGRAL    Likelihood = -0.85   Transmembrane    205 - 221 ( 205 - 221)
    INTEGRAL    Likelihood = -0.59   Transmembrane    171 - 187 ( 171 - 187)
    INTEGRAL    Likelihood = -0.53   Transmembrane    226 - 242 ( 226 - 242)

----- Final Results -----
40  bacterial membrane --- Certainty=0.1341(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

45 >CP:CB1345:GB:Z99112 similar to b-serine dehydratase [Bacillus subtilis]
Identities = 176/289 (60%), Positives = 224/289 (76%), Gaps = 1/289 (0%)

Query: 1 MPTTTELLVEQANSQKGNIAELMIQTEILMVEQNSREIRYDSNRNLVMAKSVIDGLTP 60
MF ++H+E + + I++HMI E+E+T ++S+I M NL VMA+V V GL
Sbjct: 1 MERNVKELE-ITREKQILSDMIQAEEMVEITKQSDIFQCDNHLNLSMAAQQVQKLE 59

50 Query: 61 SKSISITGSGDAVMDCQYLQSGKTIIDTTLIAAVRNAMVAINELNMGLCVATPTPSAG 120
S +GLTGSGDAV+ Y+S+SGK+S IL AV A+A N+NA MG +CATPTPSAG
Sbjct: 60 VISGTGSGDAVKLQAYTRSGKSGELLGLIDVSKAVTNENVAAGTICATPTPSAG 119

-1147-

Query: 121 CLPAVISTAIKILMLTSEKQLDFLPTAGAGFLVIGNNASISGAGGCCQAEVGSASAMAAA 180
 +P + EKLIV T E+ + FLPTAGAGF V+ NNASISGA GCCQAEVGSAS MAAA
 Sbjct: 120 VVPGTLEFAVKEKILMPTRQMIKFLPTAGAGFVFNANASISGAGGCCQAEVGSASGMAAA 179

5 Query: 181 ALVMAAGGTTPFQASQAIAFVIONMLGLICDPVAGLVEVPCVKRNALGSSFALVAADMLA 240
 A+V AGGT Q++A+A +K+NLGL+CDPVAGLVEVPCVKRNA+G+S A++AADMLA
 Sbjct: 180 AIVMAGGTTPQASRAMAITLKNMLGLVCDPVAGLVEVPCVKRNAMASNAIAADMLA 239

10 Query: 241 GIESQIPVDEVIDAMYQVGSSELPFARETAREGLAATPTGRYSKEIFG 289
 GI S+IP DEVIDAMY++G ++PTA RET +GGLAATPTGR K+IFG
 Sbjct: 240 GITSRIPCDEVIDAMYKIQGTMTALRETQGGGLAATPTGRELEKIKIPG 288

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3185> which encodes the amino acid sequence <SEQ ID 3186>. Analysis of this protein sequence reveals the following:

15 Possible site: 55
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.12 Transmembrane 196 - 212 (196 - 213)
 INTEGRAL Likelihood = -0.27 Transmembrane 226 - 242 (226 - 242)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.1447 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

25 The protein has homology with the following sequences in the databases:

>GP: CAB13459 GB: Z99112 similar to L-serine dehydratase [Bacillus subtilis]
 Identities = 173/289 (59%), Positives = 222/289 (75%), Gaps = 1/289 (0%)

30 Query: 1 MFTTIELVKQADQFPNGNIARMLIATEVEHSGRNREDI IKMSRLNQVMAAETEGITS 60
 MF ++EL++ ++ I++MIA E+E++ + EDI + M NL VM+AV +GL
 Sbjct: 1 MFRNVKELIEITKEK-QILISDVMLAQEMEVEKTEKEDIPQQMDHNLVMEAAVQRIEIG 59

35 Query: 61 TKSISGLTGGDAVIDMYIKKNSLSDTTILAVRNALAVNELNAGMLVCATPTAGSAG 120
 S +GLTGGDAVK+ VI+ G SLS IL+AV A+ A NE+NA MG +CATPTAGSAG
 Sbjct: 60 VTSQGLTGGDAVKLQVIRSGKSLSGPLIDAVSKAVATNEVNAAMTTCATPTAGSAG 119

40 Query: 121 CLPAVISTAIKILMLTSEKQLDFLPTAGAGFLVIGNNASISGAGGCCQAEVGSASAMAAA 180
 +P L EKL+ + ++ FLPTAGAGF V+ NNASISGA GCCQAEVGSAS M+AA
 Sbjct: 120 VVPGTLEFAVKEKILMPTRQMIKFLPTAGAGFVFNANASISGAGGCCQAEVGSASGMAAA 179

45 Query: 181 ALVMAAGGTTPFQASQAIAFVIONMLGLICDPVAGLVEVPCVKRNALGSSFALVAADMLA 240
 A+V+ AGGT Q++A+A +K+NLGLVCDPVAGLVEVPCVKRNA+GAS A++AADMLA
 Sbjct: 180 AIVMAGGTTPQASRAMAITLKNMLGLVCDPVAGLVEVPCVKRNAMASNAIAADMLA 239

45 Query: 241 DIDSQIPVDEVIDAMYQVGSAMPTAFRETAREGLAATPTGRYSKEIFG 289
 I S+IP DEVIDAMY++G MPTA RET +GGLAATPTGR +IFG
 Sbjct: 240 GITSRIPCDEVIDAMYKIQGTMTALRETQGGGLAATPTGRELEKIKIPG 288

An alignment of the GAS and GBS proteins is shown below.

50 Identities = 244/290 (84%), Positives = 273/290 (94%)

Query: 1 MFTTIELVKQADQFPNGNIARMLIATEVEHSGRNREDI IKMSRLNQVMAAETEGITS 60
 MFTTIELV+QA+ Q GNIAELMI TE+EM+GR+EA+I IMSRNL+VMKA+V +GLT
 Sbjct: 1 MFTTIELVKQADQFPNGNIARMLIATEVEHSGRNREDI IKMSRLNQVMAAETEGITS 60

55 Query: 61 TKSISGLTGGDAVIDMYIKKNSLSDTTILAVRNALAVNELNAGMLVCATPTAGSAG 120
 +KSISGLTGGDAVIDM Y+ G +SDDTIL AVRNA+AVNELNAGMLVCATPTAGSAG
 Sbjct: 61 TKSISGLTGGDAVIDMYIKKNSLSDTTILAVRNALAVNELNAGMLVCATPTAGSAG 120

60 Query: 121 CLPAVISTAIKILMLTSEKQLDFLPTAGAGFLVIGNNASISGAGGCCQAEVGSASAMAAA 180
 CLPAV++TAIKGL+L+E+QGL+FLPTAGAGFLVIGNNASISGAGGCCQAEVGSASAMAAA
 Sbjct: 121 CLPAVLATAIKILMLTSEKQLDFLPTAGAGFLVIGNNASISGAGGCCQAEVGSASAMAAA 180

Query: 181 ALVMAAGGTTPFQASQAIAFVIONMLGLICDPVAGLVEVPCVKRNALGSSFALVAADMLA 240

-1148-

ALV AAGGT QASQAIAFVIRN+LGL+CDPVAGLVEVPCVKRNALG+SFALVAADMALA
 Sbjct: 181 ALVKAAGTISHQASQAIAFVIRNLLGLVCDPVAGLVEVPCVKRNALGASFAVAADMALA 240

Query: 241 GIESQIPVDEVIDAMYQVGSLLPTAFRETAGGLAATPTGRYSKETINGE 290
 I+SQIPVDEVIDAMYQVGS+PTAFRETAGGLAATPTGRYS EINGE
 Sbjct: 241 DIDSQIPVDEVIDAMYQVGSAMPTAFRETAGGLAATPTGRYSVEIKGE 290

SEQ ID 3184 (GBS358) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 176 (lane 6; MW 35kDa).

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1032

A DNA sequence (GBSx1104) was identified in *S. agalactiae* <SEQ ID 3187> which encodes the amino acid sequence <SEQ ID 3188>. Analysis of this protein sequence reveals the following:

- 15 Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.
- Final Results -----
 20 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 25 >GP:BA06216 GB:AP001515 L-serine dehydratase beta subunit [Bacillus halodurans]
 Identities = 101/216 (46%), Positives = 156/216 (71%), Gaps = 2/216 (0%)
- Query: 4 LKFGVFDIIGPVMIGPSSSHPTAGAVRIGKVHVSIPGE-PSEVIFHLYNSPAKTYQGHGT 62
 +K++VVDIIGPVMIGPSSSHPTAGA RIG+V ++FG+ P + Y SFA+TY+QGHGT
 Sbjct: 1 MKYRTVFDIIGPVMIGPSSSHPTAGARIGRVARTLFGQGFRCDIYFGSFASTYKHSHT 60
- 30 Query: 63 DKALVAGILGNDTINFDIKNSLEIAHQGIKITYMDILKDSNSHENTAKITVKQKDRMS 122
 D A+V GIL DT +P I SL++A +KG++Y+ +++ + HENTAK+ ++ G+ +
 Sbjct: 61 DVAIVGILDFTDFDPRIPRLQLAKEKGVVVYFHE-HEALTDHNTAKRVVLQKGEDQLE 119
- 35 Query: 123 IFQVSIQGGNIQVTEINLNGFSVSLTMTFTLITVHQDIPGMIAKVTDLSDFNINIAQNV 182
 +GVSIQGG I++ ELANGF + L+ N P +++VH D G+IA V++L+ INI N V
 Sbjct: 120 VVGVSIGGGKTIETVSLNGFHLKLGNGHPAILVHTDRFGVIASVSNMLAKHEINIGHMEV 179
- 40 Query: 183 TRESAGEKAIMITEVDSDRCQQAQVKKIEALPHLHW 218
 +R+ G++A+N+IEVD ++++E +P++ V
 Sbjct: 180 SRKEKGKALMWIEVDQNVDDLLQLERLNPITV 215

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3189> which encodes the amino acid sequence <SEQ ID 3190>. Analysis of this protein sequence reveals the following:

- 45 Possible site: 30
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
 50 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9161> which encodes the amino acid sequence <SEQ ID 9162>. Analysis of this protein sequence reveals the following:

- 55 Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty= 0.300(Affirmative) < succ>
 bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 187/223 (83%), Positives = 205/223 (91%), Gaps = 1/223 (0%)

Query: 1 MKHLKFQSVFDIIGPVMIGPSSSHTAGAVRIGKVHVSIFG-DEVTIPLYNSFARTYQ 59
 M KFGSVFDIIGPVMIGPSSSHTAGAVRIGKVHVSIFG- P EVTIPLYNSFARTY+G
 Sbjct: 3 MNTQKFGSVFDIIGPVMIGPSSSHTAGAVRIGKVHVSIFGDIPEVTIPLYNSFARTYR 62

Query: 60 HGTDKALVAGILGMDTNDPDKNSLEIAHQKGIKIYWDILKDSNSPHNTAKITVNGDR 119
 HGTDKALVAGI+GM TNDPDKNSLEIAHQKGIKIYWDILKDSN+PHNPT KI+VK D+
 Sbjct: 63 HGTDKALVAGIMGMDTNDPDKNSLEIAHQKGIKIYWDILKDSNAPHNPTVKISVKKADK 122

Query: 120 SMSITGVSIQGGNIQVTELNQFSVSLTMTPTTLIIHVQDIPGMIAKVTDLISDFNINIAQ 179
 ++S+TGVSIQGGNIQVTELNQFSVSL+MNTPT++ VH+DIPGMIAKVTDLIS NINIA
 Sbjct: 123 TLSVTVGSIQGGNIQVTELNQFSVSLMNTPTTIVTVHKDIPGMIAKVTDLISNNINIA 182

Query: 180 MNTVRESAGEKAIMIEVDSRDCQAVKKIEAIPHILHNVNFFD 222
 MNTVRESAGEKA MIEVDSR+Q+Q+A +I IPH+HNVNFFD
 Sbjct: 183 MNTVRESAGEKAIMIEVDSRDCQAVKKIEAIPHILHNVNFFD 225

SEQ ID 3188 (GBS151) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 3; MW 50kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 188 (lane 11; MW 25kDa) and in Figure 165 (lane 14-16; MW 25.3kDa).

The GBS151-GST fusion product was purified (Figure 198, lane 3; Figure 236, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 289), which confirmed that the protein is immunoreactive on GBS bacteria.

GBS151L was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 127 (lane 8-10; MW 50kDa). GBS151L was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 127 (lane 11 & 12; MW 25kDa), in Figure 128 (lane 7; MW 25kDa) and in Figure 180 (lane 7; MW 25kDa). Purified GBS151L-His is shown in Figure 232 (lanes 5 & 6) and in Figure 240 (lanes 3 & 4).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1033

A DNA sequence (GBSx1105) was identified in *S.agalactiae* <SEQ ID 3191> which encodes the amino acid sequence <SEQ ID 3192>. This protein is predicted to be tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase (trmU). Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2208(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10291> which encodes amino acid sequence <SEQ ID 10292> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5 >GP:BA04980 GB:AP001511
   (5-methylaminomethyl-2-thiouridylyl)-methyltransferase
   [Bacillus halodurans]
   Identities = 250/359 (69%), Positives = 292/359 (80%), Gaps = 6/359 (1%)

10 Query: 32 RVVVGMSGGVDSVIALLLKEQGVYVIGVPMKNWDDTDFGVCTATEDYKDVAAVADIG 91
   RVVVGMSGGVDSVIALLLKEQGVYVIGVPMKNWDDTDFGVCTATEDYKDVAAVADIG 91
   Sbjct: 10 RVVVGMSGGVDSVIALLLKEQGVYVIGVPMKNWDDTDFGVCTATEDYKDVVQVQNLG 69

   Query: 92 IPYYSVNFPEKEYNDVFEYFLAEYRAGRITNPDVMCNKEIKFKAFLOYAMTLGADYVATG 151
   I YY+VNFPEKEYND+VF YFL EY+AGRTNPDVMCNKEIKFKAFLO+AA+TLGADYVATG
15 Sbjct: 70 IAYAVNFPEKEYNDKVFTYFLAEYKAGRTNPDVMCNKEIKFKAFLOHALTLGADYVATG 129

   Query: 152 HYAQVTRDENGIVHMLRGADNKKDQTYFLSQLSQRLQKTLFPLGHLQKPEVRIAEAG 211
   HYAQV ++ +G ++RG D KKQTYFL+ LSQ+QL + +FPLGHL+K EVR IAE AG
20 Sbjct: 130 HYAQV-KNVDDQYQLIRGDKPKNDQTYFLNALSQQLSRVMFPLGHLKKEVRIAEAG 188

   Query: 212 LATAKKDSTGICFIGEYKFKDPLQQLPLPQGRMTVDGDMGEHAGLMYTTIGQRGL 271
   LATAKKDSTGICFIG+ ++FK+FL YLPAPQG M T+DG G H GLMYTT+GQR GL
25 Sbjct: 189 LATAKKDSTGICFIGRDKPKFLSSYLPAPGEMQTLDEGVKGTGHDGLMYTTIGQRGL 248

   Query: 272 GIGGSHGSDNKPWFVVGKDLKSNILVVGQFVHDSLMSTGLTASEIHPTFRDMNPEKLEC 331
   GI GG ++PWFV+GK+L KNILVVGQF+H L S L A +++ ++ EC
30 Sbjct: 249 GI----GSGSEPFVVGKDLKSNILVVGQFHPGLYSEGLRAIKVNMILRRSEDPPEC 304

   Query: 332 TAKFRYRQPDKVTYVYVVGKQA-RVVFDELQRAITPQAVVFNBEQCLGGMIDQAYR 389
   TAKFRYRQPD KVTYVY ++ A V+V + QRAITPQAVVFN+ CLGGG ID +
35 Sbjct: 305 TAKFRYRQPDKVTYVYVVGKQAVGVLPFAEQRAITPQAVVFNVDGVCLGGGTIDRVLK 363

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3193> which encodes the amino acid sequence <SEQ ID 3194>. Analysis of this protein sequence reveals the following:

```

35 Possible site: 29
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
40 bacterial cytoplasm --- Certainty=0.1691 (Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

   RGD motif: 331-333

```

The protein has homology with the following sequences in the databases:

```

45 >GP:BA04980 GB:AP001511
   (5-methylaminomethyl-2-thiouridylyl)-methyltransferase
   [Bacillus halodurans]
   Identities = 255/359 (71%), Positives = 293/359 (81%), Gaps = 6/359 (1%)

50 Query: 14 RVVVGMSGGVDSVIALLLKEQGVYVIGVPMKNWDDTDFGVCTATEDYKDVAAVADIG 73
   RVVVGMSGGVDSVIALLLKEQGVYVIGVPMKNWDDTDFGVCTATEDYKDVAAVADIG 73
   Sbjct: 10 RVVVGMSGGVDSVIALLLKEQGVYVIGVPMKNWDDTDFGVCTATEDYKDVVQVQNLG 69

   Query: 74 IPYYSVNFPEKEYNDVFEYFLAEYRAGRITNPDVMCNKEIKFKAFLOYAMTLGADYVATG 133
   I YY+VNFPEKEYND+VF YFL EY+AGRTNPDVMCNKEIKFKAFLO+AA+TLGADYVATG
55 Sbjct: 70 IAYAVNFPEKEYNDKVFTYFLAEYKAGRTNPDVMCNKEIKFKAFLOHALTLGADYVATG 129

   Query: 134 HYAQVKTRDENGIVHMLRGADNKKDQTYFLSQLSQRLQKTLFPLGHLQKPEVRIAEAG 193
   HYAQVK ++ +G ++RG D KKQTYFL+ LSQ+QL + +FPLGHL+K EVR IAEAG
60 Sbjct: 130 HYAQVK-KNVDDQYQLIRGDKPKNDQTYFLNALSQQLSRVMFPLGHLKKEVRIAEAG 188

```


-1151-

Query: 194 LATAKKKSTGICFIGEKNFKQLSQYLAQKGRMMTIDGRDMGHAGLIMYYTIGQRGGL 253
 LATAKKKSTGICFIG++FK+PLS YLPAQ G M T+DG G H GIMYTT+QGR GL
 Sbjct: 189 LATAKKKSTGICFIGKRDFKEFLSSYLAQPGEMQTLDEGVKPHIDGLIMYYTIGQRGGL 248

5 Query: 254 GIGGCGGNDQPNWVVGKDLNQNLIVGQGFYHIALMSNSLDASVIHPTREMPPEFTPEC 313
 GI GG +PWFV+GK+L +NLLVVGQGF+H L S L A +++ + FEC
 Sbjct: 249 GI----GSGGEPWVFIGNKLHNLIVGQGFHHPGLYSEGLRAIKNVMILRRESDEFTPEC 304

10 Query: 314 TAKFRYRQDSHVAVHVRGDKA-EVVFPAEQRAITPGQAVVFDGKECLGGGIMDAYK 371
 TAKFRYRQD V V+ + D A EV+FAEQRAITPGQAVVFDG CLGGG ID K
 Sbjct: 305 TAKFRYRQDQKVTYVPSDGADEVLPFAEQRAITPGQAVVFDGVDCLGGGIDHVLK 363

An alignment of the GAS and GBS proteins is shown below.

Identities = 332/377 (88%), Positives = 349/377 (92%)

15 Query: 21 GRILGIDNSNIRVVVGMGGVDSSVTALLKEQGYDVGIVPMQWDDTDFPGVCTATEDY 80
 G MDSN IRVVVGMGGVDSSVTALLKEQGYDVGIVPMQWDDTDFPGVCTATEDY
 Sbjct: 3 GEFFMIDNSKIRVVVGMGGVDSSVTALLKEQGYDVGIVPMQWDDTDFPGVCTATEDY 62

20 Query: 81 KDVAAVADQIGIPYYSVNFKEKYNDRVFEYFLAETRAGRITFNDVCMCKEIKFKAPLOYA 140
 KDVAAVAD+IGIPYYSVNFKEKYNDRVFEYFLAETRAGRITFNDVCMCKEIKFKAPLOYA
 Sbjct: 63 KDVAAVADKIGIPYYSVNFKEKYNDRVFEYFLAETRAGRITFNDVCMCKEIKFKAPLOYA 122

25 Query: 141 MTLGADYVATGHYAQVTRDENGIVHMLRGADNNKDTYFLSQLSQEQLOKTLFPLGHLQK 200
 MTLGADYVATGHYAQV RDENG VHMRLGADN KDQTYFLSQLSQEQLOKTLFPLGHLQK
 Sbjct: 123 MTLGADYVATGHYAQVTRDENSTVHMLRGADNGKDQTYFLSQLSQEQLOKTLFPLGHLQK 182

30 Query: 201 PEVRRIASEAGLATAKKKSTGICFIGEKNFKQLSQYLAQKGRMMTIDGRDMGHAGL 260
 EVR IAR AGLATAKKKSTGICFIGEKNFK FL QYLPQA GRMMT-IDGRDMGHAGL
 Sbjct: 183 SEVRRIASEAGLATAKKKSTGICFIGEKNFKQLSQYLAQKGRMMTIDGRDMGHAGL 242

35 Query: 261 MYTTIGQRGGLGIGGCGGNDQPNWVVGKDLNQNLIVGQGFYHIALMSNSLDASVIHPT 320
 MYTTIGQRGGLGIGGCGGNDQPNWVVGKDLNQNLIVGQGFYHIALMSNSLDASVIHPT
 Sbjct: 243 MYTTIGQRGGLGIGGCGGNDQPNWVVGKDLNQNLIVGQGFYHIALMSNSLDASVIHPT 302

40 Query: 321 RMPNEFLEKCTAKFRYRQDPSKVTVVVGKQARVVFDDLQRAITPGQAVVFDGKECLG 380
 R+MP EF ECTAKFRYRQDPS V V+V+G+AA VVF +QRAITPGQAVVFDG+ECCLG
 Sbjct: 303 RMPPEFTFECTAKFRYRQDPSHVAVHVRGDKAEVVFPAEQRAITPGQAVVFDGKECLG 362

45 Query: 381 GGMIDQAYRDDKICQYI 397
 GGMID AY++ + CQYI
 Sbjct: 363 GGMIDMAYRNGQPCQYI 379

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1034

A DNA sequence (GBSx1106) was identified in *S. agalactiae* <SEQ ID 3195> which encodes the amino acid sequence <SEQ ID 3196>. Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -12.84	Transmembrane	141 - 157 (134 - 165)
INTEGRAL	Likelihood = -11.78	Transmembrane	40 - 56 (36 - 73)
INTEGRAL	Likelihood = -4.35	Transmembrane	68 - 84 (65 - 86)
INTEGRAL	Likelihood = -3.50	Transmembrane	180 - 196 (175 - 199)

55 ----- Final Results -----
 bacterial membrane --- Certainty=0.6137 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

60

The protein has homology with the following sequences in the GENPEPT database.

-1153-

[illegible]

10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1035

A DNA sequence (GBSxI107) was identified in *S.galactiae* <SEQ ID 3197> which encodes the amino acid sequence <SEQ ID 3198>. Analysis of this protein sequence reveals the following:

```

15   Possible site: 17
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
20   bacterial cytoplasm --- Certainty=0.1747(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10289> which encodes amino acid sequence <SEQ ID 10290> was also identified.

25 The protein has homology with the following sequences in the GENPEPT database.

>GP: AAC45494 GB:U80409 glucose inhibited division protein homolog
GidA [Lactococcus lactis subsp. cremoris]
Identities = 394/524 (75%), Positives = 458/524 (87%), Gaps = 2/524 (0%)

30 Query: 13 KTLATINLEMLAFMPCNPISIGSGAKGIVVREIDALGGEMGNIDKTYIQMKMLNTGKGP 72
KTLT TINL M-AFMPCNPISIGSGAKGIVVREIDALGGEMG+NIDKTYIQMKMLNTGKGP
Subject: 12 KTLMTINLNMVAFMPCNPISIGSGAKGIVVREIDALGGEMGNIDKTYIQMKMLNTGKGP 71

35 Query: 73 AVRALRAQADKALYAQIMKQTVKQENLTLRQAMIDEILVEDGK--VVGVRTATNQKFSFA 130
AVRALRAQADK YA +MK TV QENLTLRQ M++E++++D K V+GVRT+T ++ A
Subject: 72 AVRALRAQADKDEYAASMKQNTVSDQENLTLRQGVVEELILDDEKQKVGVRTSTGTQYGA 131

Query: 131 KSVVITGTALRGEIILGDLKYSSGPNNSLAVTLADNLRDLGLEIGRPKTGTTPPRVIGAS 190
K+V+ITGTALRGEII+G+LKYSSGPNNSL+S+ LADNLR++G EIGRPKTGTTPPRV AS
40 sbjct: 132 KAVIITGTALRGEIILGELKYSSGPNNSLSIGLADNLRIGEIFEIGRPKTGTTPPRVLAS 191

Query: 191 S I N Y E K T E I Q P G D E Q N H F S F M S R D E D Y I T D Q V P C W L T Y T N T L S H D I I L Q N L H R A P M F S G 250
SI+Y+KTEIQGDE PNHFSFMS DEDY+ DQ+PCWLTYT SH I+ NLHRAP+FSG
Sbjct: 192 S I D Y K T E I Q P G D E A P N H F S F M S S D E D Y L K D I P C W L T Y T T E N S H T I L R D N L H R A P L F S G 251

45
Query: 251 IVKGVGPRYCPISIEDKIVRFADKRIHQLFLEPEGRYTEEVYVQGLSTSLPEDVQVDLLRS 310
IVKGVGPRYCPISIEDKI RFADK RHOLFLEPEGR TEEVY+ GLSTS+PEDVQ DL++S

50 Query: 311 IKGLENAEMMRGTGYAIEYDIVLPQLRATLETAKVIAGLFTAGCTNGTSGYERAAAGQGLVA 370
I GLENA+MMR GYAIEYD+V+PQLR TLETK+I+GLFTAGCTNGTSGYERAAAGQGLVA
Sbjct: 312 IPGLENAKMMPRGYAIEYDVVMPQLRPTLETKLISGLFTAGCTNGTSGYERAAAGQGLVA 371

55 Query: 371 GINAALKVQGKPEILIKRSDAYIGVMIDDLVTGKLTLEPYRLTTSRAEYRILRHNDADMR 430
GINAALK+QGKPE ILKRS+AYIGVMIDDLVTGKLTLEPYRLTTSRAEYRILRHNDAD R
sbjct: 372 GINAALKIQGKPEFILKRSEAYIGVMIDDLVTGKLTLEPYRLTTSRAEYRILRHNDADRR 431

Query: 431 LFEIGYEIGLVDBERYAIFKKRQMQFENELRLDSIKLPVSETNKRIQELGFKPLTDAL 490
LFEIG ++GLV ++ ++ + QF+ E++RL+S KLKP+ +T +++ +LGF P+ DAL
60
Sbjct: 432 LFEIGRQVGLSDVAQWEHYQAKMAQFDREMKRLNSEKLKPLDPTQELGKLGKLGPGPINDAL 491

-1154-

Query: 491 TAKEFMRFPQITYAVATDFVGCADFLDSKVIELLETRIKYEGSY 534
 T EF+RP++ Y DF+G A E +D V EL+ETEI YEGSY
 Sbjct: 492 TGAZFLKRFVNVYDEVIDFIGQAFVIDRITVSELLITRITYEGSY 535

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3199> which encodes the amino acid sequence <SEQ ID 3200>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1064 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 530/610 (86%), Positives = 574/610 (93%)

Query: 1 MEASLAASRMGCKTLLATINLEMLAFMPCNPISIGSAGIVUREIDALGGEMGNIDKTY 60
 +EASLA SRMGCKTLLATINLEMLAFMPCNPISIGSAGIVUREIDALGGEMGNIDKTY
 Sbjct: 21 VEASLATSRMGCKTLLATINLEMLAFMPCNPISIGSAGIVUREIDALGGEMGNIDKTY 80

Query: 61 IQGGLATGKGPAVRLAQAQADKALYA+ QTMKTVKEQRLTLRQNMIDRILVEDGKVVGV 120
 IQGGLATGKGPAVRLAQAQADK+LYA+ MK TVEKQ NLTLRQ MID+ILVEDG+VGVV
 Sbjct: 61 IQGGLATGKGPAVRLAQAQADKSLYAREMKHTVEQANLTLRQNMIDRILVEDGKVVGV 140

Query: 121 RTATNQKFSFAKSVVITTTGTLRGEIILGDLKYSSGPNHSLASVTADNLRDLGLEIGRPF 200
 +TAT QKF+AK+VV+TTGTALRGEIIL+LKYSSGPNHSLASVTADNLR+ LGLEIGRPF
 Sbjct: 141 LTAATGQKFAAKAVVITTTGTLRGEIILGDLKYSSGPNHSLASVTADNLRDLGLEIGRPF 280

Query: 181 TGTTPRVKASSINYEKTEIQGDEQFNHPSFMSRDEYITDQVPCMLTYTNLTSHDIINQ 240
 TGTTPRVKASSIN+Y+TEIQGD++FNHPSFMS+D DY+ DQ+PCMLTYTN SHDIINQ
 Sbjct: 201 TGTTPRVKASSINVDQTEIQGDDKFNHPSFMSKADADYLDQ+PCMLTYTNLTSHDIINQ 260

Query: 241 NLARAPMPSGIVKGVPYRCPSEDRIKIVRFADKERHQLFLEPEGRYTEVYVQGLSTSLP 300
 NL+RAPMPSGIVKGVPYRCPSEDRIKIVRFADKERHQLFLEPEGR TEEVYVQGLSTSLP
 Sbjct: 261 NLARAPMPSGIVKGVPYRCPSEDRIKIVRFADKERHQLFLEPEGRDTEVYVQGLSTSLP 320

Query: 301 EDVQVDLLRSIKGLENAEMMRGTGVAIEYDVLPHQLRATLET+VAGLFTAGQNGTSGY 360
 EDVQ DL+ SIKGLE AEMMRGTGVAIEYDVLPHQLRATLET+I+GLFTAGQNGTSGY
 Sbjct: 321 EDVQVDLLRSIKGLENAEMMRGTGVAIEYDVLPHQLRATLET+VAGLFTAGQNGTSGY 380

Query: 361 EEAAGQQLVAGINAAKVKQKPELILKRSDAYIGVMIDDLVTKGTLEPYRLLTSAEYRL 420
 EEAAGQQLVAGINAAKVKQKPELILKRSDAYIGVMIDDLVTKGTLEPYRLLTSAEYRL
 Sbjct: 381 EEAAGQQLVAGINAAKVKQKPELILKRSDAYIGVMIDDLVTKGTLEPYRLLTSAEYRL 440

Query: 421 ILRHDAEMRLTEIGVEIGLVDERKVAIPKRRQMPFNELERLDSIKLKPVSRTNKRQ 480
 ILRHDAEMRLTEIG+IGLV+ER+ P+ ++ QP+NL+RL+SIKLP+ ETN R+Q+
 Sbjct: 441 ILRHDAEMRLTEIGRDIGLVDERKVAIPKRRQMPFNELERLDSIKLKPVSRTNKRQ 500

Query: 481 LGPKFLTDATAKEFMRFPQITYAVATDFVGCADFLDSKVIELLETRIKYEGSY+KALD 540
 LGPKFLTDATAKEFMRFPQITYAVATDFVGCADFLDSKVIELLETRIKYEGSY+KALD
 Sbjct: 501 LGPKFLTDATAKEFMRFPQITYAVATDFVGCADFLDSKVIELLETRIKYEGSY+KALD 560

Query: 541 QVAKMKRMEEKRIPIPHIDWDIDSIAETARQKFKKINPETAQASRISGVNPAISILM+ 600
 QVAKMKRMEEKRIPIPHIDWDIDSIAETARQKFKKINPETAQASRISGVNPAISILM+
 Sbjct: 561 QVAKMKRMEEKRIPIPHIDWDIDSIAETARQKFKKINPETAQASRISGVNPAISILMI 620

Query: 601 YLEGRGKGRK 610
 YLEGR K +
 Sbjct: 621 YLEGRGKAER 630

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1155-

Example 1036

A DNA sequence (GBSx1108) was identified in *S. agalactiae* <SEQ ID 3201> which encodes the amino acid sequence <SEQ ID 3202>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have a cleavable N-term signal seq.
 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB07750 GB:AP001520 unknown conserved protein in B. subtilis
 [Bacillus halodurans]
 15 Identities = 205/644 (31%), Positives = 362/644 (55%), Gaps = 28/644 (4%)
 Query: 35 LLLAIFVALSFVVALLYQ-----KITYELSEVBQIELNDQTE 73
 ++ + VAL P++AL +YO +I++E + I L+ +
 Sbjct: 14 VIALLAVALPLIALSFYQMLGVIGVLLLVIAIFSLRARISPERDLEQYISTLEYRVH 73
 Query: 74 VSLKBLBQMPGVQVQFDLENDIEWFPYA-ELIPTDNGHFQSATVVDIITSRNGTA 132
 + + + Q+PVG+I ++ + ++M NPYA E + + +++ + GT
 Sbjct: 74 KAGEKAVTQLPVGMLIYNDQLR-VQWVNPYAASHLPKAEINASLELSELPVLALESGTD 132
 Query: 133 GQSFPEYDKNKYSAVLOTETGVFFFDNFMNRRNDGSMLEFVIGLISIDNYDDIMDML 192
 Q + Y + YFD R + +PV+ I +DNYD++ M
 Sbjct: 133 EQKIVIEEKTYDCTFKPNERLIYFFDITSESRMHQPFESQPVLTFTIYLDNDEVITQME 192
 Query: 193 EADMSKINAVTFSFISDFTQKNI FYRRVNMMDRYIFTDYSLVNLIKDKFDILNEFKR 252
 + S++ + VTS ++ + ++F RR DR+ Y L + K KF IL+R R+
 Sbjct: 193 DQVRSRLMSQVTSLSLQWNEHDLFLRTAADRFLVMSYGSLLAIEKTFGLIDIEIRET 252
 Query: 253 AQENHLSLTSMGISTYGDGNHNGIGQIALENLNTALVGGGDQIVVRENDSSKALYFGG 312
 + + I+LS+G+ YGD + ++GQ+A +L+ AL RGGDQ+ +++ K ++G
 Sbjct: 253 TGKREKIP+LTSIGVGTGDLGLRELGLQAQSSLDLALRGGDQVAIKQRT- KRVYGGK 310
 Query: 313 AVSTIKRSRTTRAMWTAISDLKVVDSVPVIGHRKLMDALASVGMOQFASINVASY 372
 + + KR+R R R + A+ D + D V ++GH+ DMDA+GA++G+ A ++
 Sbjct: 311 SNAMKRTKRVARVISHALEDFVLESDRVIMVGHNPMDAVGAIGILKIAEVNDRAF 370
 Query: 373 VVYDNDMNSDIERAIDYLDQGET- -RLVSVERAFELITQNSLVAMVDHSKITALTSLKE 430
 VV DND+N D+ + + + + + + + + E + EL+T+ +LLV+VD K + + +
 Sbjct: 371 VVLDNDVNDPVS KMEVEKNEQLWMDKFTIPESLELMTSETLIVDTKPSMVIEPR 430
 Query: 431 FPNKPADVIVVDHHRDEDFPNKAVLSFIEGASSASELVTELQPOQKDKLSRQASI 490
 + V+V+DHRR E+F ++ VL ++E ASS +ELVEL+++Q K K+ + + +
 Sbjct: 431 LLVDYVERVVVDHHRREGFIEDPVLVYMERPYASSTAEVLTELLEYPQKLKMDILESTA 490
 Query: 491 IMAGIMLDTNRNFASNVTSRTFDVASYLRLGSGNSMAIQKISATDFEYKRLNELILKQR 550
 L+AG++H++FA +RFD AS+LR G+++ +QK+ D + Y +L+ + + +
 Sbjct: 491 LLAQMVIVDTKSFARVIGARTFVAASFLRSHGADTVLVQKLKELDHYVTKRAKIVETAKL 550
 Query: 551 IYDNIIVATGEEHKVYSHVIASKAADTMLIMAGIEATFVITKNSSN-IGISARSNNIN 609
 D + +AT E + S ++ +AADT+LTM G+ A+FVI++ + ISARS ++NV
 Sbjct: 551 YRDGMAITAREEEAVGQLLAQADTLITMKGVVASFVIGRRHGDVVISIARSLGVNV 610
 Query: 610 QRIMEKLGCGGHFSFAACQIQKSVKQVRMMLLEIIDELELRNS 653
 Q IME L GGGH + AA Q +D + + + + L E ID+ L S S
 Sbjct: 611 QLIMESLDGGHITNAATQPRDATLSERAKLKEAIDQYLEGGS 654

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3203> which encodes the amino acid sequence <SEQ ID 3204>. Analysis of this protein sequence reveals the following:

Possible site: 25

-1156-

```
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL    Likelihood =-18.57    Transmembrane    33 - 49 ( 6 - 56)
INTEGRAL    Likelihood =-10.14    Transmembrane    12 - 28 ( 6 - 32)

5  ----- Final Results -----
      bacterial membrane --- Certainty=0.8429 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

10 The protein has homology with the following sequences in the databases:

```
>GP:BAB07750 GB:AP001520 unknown conserved protein in B. subtilis
[Bacillus halodurans]
Identities = 199/659 (30%), Positives = 367/659 (55%), Gaps = 16/659 (2%)

15 Query: 1 MKKF---RFETIHLI-MNGLILFGLIALCVSIHQSKILLILAIFLVILFV-ALLMYQKE 55
      M KF R+ H+I ++ + L L+AL Q ++ +L + ++ +P + A + +++
      Sbjct: 1 MPKFLLRKRWGHYIVALLAVALVFLIALSTYQWQLGVIGVLLLVIAIFSLRARIISFERD 60

20 Query: 56 AYQLSDLAHISLINEQTEQNLKTLTLLNMPVGVVQDQTNVAVENYFPA-ELIPTTEEGF 114
      Q +I L+ + + + +PVG+ + + + V+W NPYA E + E
      Sbjct: 61 LEQ-----YISTLSYRVHKAGEEAVTLQVGMILYNDQLR-VQWVNPYAAEHLFKPEIDA 114

Query: 115 IQNGLIQIITEKRRRDISQTEFVSQNKYTSYIDVSSGIFPYFFDSFVGNRLADASMLRP 174
      L +++ Q + Y + + YFD DR+ +P
25 Sbjct: 115 SLEELSPELVKALEEGTDEQKIVIEKTYDCTCFKNERLIYFFDITSESRMQQFEEQSP 174

Query: 175 VVGIIISVDYDDITDDLSADATSKINSFVANFIDFPMESKRIPYRVNMDRYFFDFDKT 234
      V+ I +DRYD+T + D S++ S V + +++ +F RR DR+ ++
30 Sbjct: 175 VLTFTYLDNYDVTQGMEDQVRSRLMSQVTSLSNQNANHEHDLFLRTAADRPIAVMSYGS 234

Query: 235 LNDLMDNKFVSLBEFRKEAQDQRPLTSLIGISPGNEHNSIQIGVALENLALVRGGDQ 294
      L + KF +L+E R+ + PTLISIGD+ +G+ + ++QQA +L+AL RGGDQ
      Sbjct: 235 LLAIEKTKFGLIDREITTTGKEKIPLTLSIGVGDLSLRELOQLAQSSLDLALGRGGDQ 294

35 Query: 295 IVIRENADHTNPIYFGGSGSVTSVRKTRTRAMTAISDRIKMDVNPFVHKLMDAL 354
      + I++ ++GG S + KR+R R+ + A+ D + D V ++GH+ DMDA+
      Sbjct: 295 VALKQKTGKVR--FYGGKSNMKEKTRVRARVISHALRDFVLESDRVIVMGRKNPMDAV 352

40 Query: 355 GSAVMGQFPAGNIENSFAVYNPEMSPIERAIRRLQADGKT--RLISVSQAMGLVTPR 412
      G+A+G+ A +F V +P+++ED+ +E ++ + + I+ +++ L+T
      Sbjct: 353 GAAIGILKIAEVNDREAPVFLDPNDVNFVSKLIMEVEKNEQMDKFITPEESLEIATEE 412

Query: 413 SLAVMDHSKISLITLSKEFYEQFNVIVDHHRDDDFPQNAILTFIESQASSAELVTE 472
      +LIV+VD K S+ + + + V+V+DHHR ++F ++ +L ++E ASS AELVTE
45 Sbjct: 413 TLLVIDVTHKPSMVIEPRLLDYVERVVLVDHHRGEEFIEDPVLVYMEFYASSTAEVTE 472

Query: 473 LIQFNAAKKCLNKIQASVLAMIGIMLOTNPFSTRVTSRTPDVASVLRSGSDSVEIQNISA 532
      L+++Q K ++ +++ L+AG+++DTK+P+ R +ETPD AS+LRS G+D+V + Q+ +
      Sbjct: 473 LLEYQPKKLKMDILESTALLAGMIVDTKSPAIRTGARTPDAAFLRSHGADTVLQKLLK 532

50 Query: 533 TDFEYKQINETILQGERIGDSIIVAAGEKNHLYSNVIAKARDTILSMARHVASPFLVE 592
      D Y + +++ + D +A + S ++ ++ADT+L+M V ASPV+
      Sbjct: 533 EULNHVYKRAVLVETAKLYRDGMATATAREEAVSQLLIQAQADTLITMGVVASPVSIR 592

55 Query: 593 TASHKIAISARSRSKINQVRVMEKLGQGGHFNLAACOLTDISLQAKYILLKTTINTMK 651
      ++ISAKS +NVQ +ME L GQGH AA Q D +L ++ + L + I+ ++
      Sbjct: 593 RHDGVSTISARSLGVNVQLIMEGLDGGHITNAATQPEDATLEAEAKVLRKADQYLE 651
```

An alignment of the GAS and GBS proteins is shown below.

```
60 Identities = 428/658 (65%), Positives = 547/658 (83%), Gaps = 1/658 (0%)

Query: 1 MKRFRFATVHLVILGLILFGLIAICVRLFQSYTALLAIFVALSFVVALLYQKITYLES 60
      MK+FRF T+HL++GLILFGLIA+CV + QS +LIAIF+ L FVVAL+YQK Y+LS
      Sbjct: 1 MKKFRFETTHLIMGLILFGLIALCVSIHQSKILLILAIFLVILFVVALVYQKRAYQLS 60

65 Query: 61 EVEQTELLNDQTEVSLKSLLEQMPGVGVIQDLETNDIENFNPYAEILFPGDNGHPOSATV 120
```

-1157-

++ IRELN+QTE +LK+LL+ MPVGV+QPD ETN +EW+NPYAEILPT + G Q+ +
 Sbjct: 61 DLAHIEELN+QTEHNLKTLLENDMPVGVVQFDQETNAVEWYNPYAEILPTTEHGIQNL 120
 5 Query: 121 KDIITSRRTAGTQGSFEYGDNKSAYLDTETGVFPFDFMGNRRNYSMLRPFVIGIS 180
 + IIT +R Q+FE NKY+Y+D +G+FYFDF+P+GMR+ D+SMLEKRPV+GIIIS
 Sbjct: 121 QQIITEKRREDSQITFEVSGNKYTSIDVSSGIFFYFDFSVGNQLADASMLRPFVIGIS 180
 Query: 181 IDNYDDIMEITMLEADMSKINAPVTSFISDFTCQKNIPYRVMNDRYYIITDYSVLNLTAK 240
 +DNYDDI D +AD SKIN+EV +FI +F +SK IFYRVMNDRYY FTD+ LN L+
 10 Sbjct: 181 VDNYYDDITDLSADTSSKINSFVANFIDFPMESKRIFPYRVMNDRYYFPTDFKTIANDMD 240
 Query: 241 DKFDILAEFRKRAQENHLSITLMSGISYGDGNHNIQGLEENLALVARGGDQIVVREN 300
 +KF +L ERK AK+ LTL+G+IS+G+ NH+QIGQ+ALENL+L ALVARGGDQIV+REN
 Sbjct: 241 NKFSVLEEFKRAQDAQRPLTLSTIGISFGESNHSQIQVALENLALVARGGDQIVIREN 300
 15 Query: 301 DSSKCALYFGGGAVSTIKRSRTTRAMMTAISDRKVDGVFTVCHHKLMDALGASVGM 360
 +YFGGG+VST+KPSRTTRAMMTAISDR+K+VD+VFTVCHHKLMDALG+AVGM
 Sbjct: 301 ADHTNPIYFGGGSVTSVKRSRTTRAMMTAISDRKVDGVFTVCHHKLMDALGASVGM 360
 20 Query: 361 QFFASINVASVYVVDPMDSNDIERAIDYLEDGFTIRLVSVERAFELITQNSLLVMVEH 420
 QFFA NI+ S+ YF+P+M+M+ DIERAI+ LQ DG+TRL+GV +A L+T SLVAVVDH
 Sbjct: 361 QFFAGNIENSFAVINPDEMSFDIERAIERLQADGKTRLSISVSQWGLVTPRSLVAVVDH 420
 25 Query: 421 SKTALTLSEKFPFKADVIIVDHHRRDEDFPKNVLFSIESGASSASELVTELIQPCQAK 480
 SK +LTLSEKF+ +F +VIVDHHRRD+DFP NA+L+FIESGASSA+SLVTELIQPCQ AK
 Sbjct: 421 SKISLTLSEKFPYEQPNVIVDHHRRDODFONAILFIESGASSAASELVTELIQPCQAK 480
 Query: 481 DKLGSSQASILMAGIMLDTNPFASINVTSTFDVASYLGLGNSMAIQKISATDPDEYRL 540
 L++ QAS+LMAGIMLDT+NE++ VTSRTFDVASYL RGS+Q+ IQ ISATDF+EY+
 30 Sbjct: 481 KCINKIQASVLAMGIMLDTKPNFSTRVTSRTFDVASYLRSKGSDSVEIQNISATDPDEYKQ 540
 Query: 541 INELILKGERIYDNIIVATGEEHVKYSHVIASKAADTLMTAGIEATFVITKNSG+IGI 599
 INE+IL+GER+ D+IIIVA GE++ +YS+VIASKAADT+L+MA +EA+FV+ + +S+ I I
 Sbjct: 541 INEILKGERLGSIIIVAAEGEHNLYSNVIAKAAATILMAHVEASFLVETASKEIAI 600
 35 Query: 600 SARSRNNINVRIMEKLGCGGSHFSAACQIDQKSVKQVRMLLITIEDRENSTVEN 657
 SARSR+ INVR+MEKLGCGGSHF+ AACQ+ D S+ Q+ +L+L+ I+ ++E VE+
 Sbjct: 601 SARSRKINVRIMEKLGCGGSHFNLAACQLDITSLPQAKYLLLTINWTKETQVEBS 658

40 A related GBS gene <SEQ ID 8717> and protein <SEQ ID 8718> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 9
 MoG: Discrim Score: 13.82
 GvA: Signal Score (-7.5): -0.890001
 45 Possible site: 44
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 0 value: 2.97 threshold: 0.0
 PERIPHERAL Likelihood = 2.97 574
 modified ALOM score: -1.09
 50 *** Reasoning Step: 3
 ----- Final Results -----
 55 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

31.3/55.8% over 631aa
 Bacillus subtilis
 60 EGAD[19304] hypothetical 74.3 kd protein in rpli-ctof intergenic region Insert
 characterized
 SP[P37484]YYBT_BACSU HYPOTHETICAL 74.3 KDA PROTEIN IN RPLI-CTOF INTERGENIC REGION. Insert
 characterized
 GP[467336][dbj]BA05182.1||D26185 unknown Insert characterized
 65 GP[2636598][emb]CAB16088.1||Z99124 yybt Insert characterized

SEQ ID 8718 (GBS10) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 1 (lane 6; MW 98kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 2 (lane 7; MW 73kDa).

The GST-fusion protein was purified as shown in Figure 189, lane 3.

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1037

A DNA sequence (GBSx1109) was identified in *S. agalactiae* <SEQ ID 3205> which encodes the amino acid sequence <SEQ ID 3206>. Analysis of this protein sequence reveals the following:

10 Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.4643 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:CAA43972 GB:X62002 ribosomal protein L9 [Bacillus
stearothermophilus]
Identities = 80/149 (53%), Positives = 105/149 (69%), Gaps = 2/149 (1%)

Query: 1 MKVIFLDVKGKGGKKEVKEVPTGYAQNFLKKNLAKETTTQAIGELKGGKQSEBEKAQAS 60
MKVIFLDVKGKGGKKEVKEVPTGYAQNFLKKNLAKETTTQAIGELKGGKQSEBEKAQAS 60

25 Sbjct: 1 MKVIFLDVKGKGGKKEVKEVPTGYAQNFLKKNLAKETTTQAIGELKGGKQSEBEKAQAS 60

Query: 61 ILAQAKELKTQLESETTRVQFIEKVGPDGRTFGSITAKKIABELQKQYGIKIDRRHIDL 120
LA AK+LK QLE T + K G GR FGSIIT+K+IAE LQ Q+G+K+DKR I+L

30 Sbjct: 61 ELANAKKLEQLKELKTTVTP--AKAGEGGRFGSITSKQIAESLQAGHLKLDKRRKIELA 118

Query: 121 HTIRAIKGVFVFKLEKQVSSQIKDLIDE 149
IRA+G VPKLEH +V++ +K+ +E

35 Sbjct: 119 DAIRALGYTNVFKLHPEVTATLKVHIVE 147

- 35 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3207> which encodes the amino acid sequence <SEQ ID 3208>. Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.4630 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 45 An alignment of the GAS and GBS proteins is shown below.

Identities = 119/150 (79%), Positives = 138/150 (91%)

Query: 1 MKVIFLDVKGKGGKKEVKEVPTGYAQNFLKKNLAKETTTQAIGELKGGKQSEBEKAQAS 60
MKVIFLDVKGKGGKKEVKEVPTGYAQNFLKKNLAKETTTQAIGELKGGKQSEBEKAQAS 60

50 Sbjct: 1 MKVIFLDVKGKGGKKEVKEVPTGYAQNFLKKNLAKETTTQAIGELKGGKQSEBEKAQAS 60

Query: 61 ILAQAKELKTQLESETTRVQFIEKVGPDGRTFGSITAKKIABELQKQYGIKIDRRHIDL 120
ILA+A+ +K L+ + TRVQF EKVGPDGRTFGSITAKKI+BELQKQ+G+K+DKRH I LD

55 Sbjct: 61 ILAQAQVAVKLDKRTKRVQFIEKVGPDGRTFGSITAKKISBELQKQFVVKVDKRRHIVLD 120

-1160-

Query: 121 HTIRAIGKVEVPVKLHKQVSSQ¹KLDIKEA 150
 H IRAIG +EVPVKLHK+V+++IKL I EA
 Sbjct: 121 HPIRAIGLIEVPVKLHKVEVTAETKLAITEA 150

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1038

- A DNA sequence (GBSx1110) was identified in *S.agalactiae* <SEQ ID 3209> which encodes the amino acid sequence <SEQ ID 3210>. This protein is predicted to be DNA polymerase III delta prime subunit (dnaB). Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -0.43 Transmembrane 204 - 220 (204 - 220)

- 15 ----- Final Results -----
 bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2423> which encodes the amino acid sequence <SEQ ID 2424>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.27 Transmembrane 210 - 226 (210 - 226)

- 25 ----- Final Results -----
 bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 30 An alignment of the GAS and GBS proteins is shown below.

Identities = 397/450 (88%), Positives = 431/450 (95%), Gaps = 1/450 (0%)

- 35 Query: 3 EYSELRVQPDLLAEQAVLGSIFISPEKLIMVREPISPDIFYKYSHKIVFRAMITLADR 62
 EV+ELRVQPDLLAEQ+VLGSIFISP+KLI VREFISPDIFYKI+HK+IFRAMITL+DRN
 Sbjct: 8 EYSELRVQPDLLAEQAVLGSIFISPEKLIAVREPISPDIFYKYAHKIFRAMITLSDRN 67
- 40 Query: 63 DAIDAMTVRNILDQDQLNGIGLGYIVELVNSVPTISANAEFYAKIVSEKAMLEDIISK 122
 DAIDA T+R ILDDQ DLQ+IGGL YIVELVNSVPTISANA+YAKIV+EKAMLEDII+L
 Sbjct: 68 DAIDATTIRILDQDQDLQIGLGYIVELVNSVPTISANAERYAKIVAEKAMLEDIARL 127
- 45 Query: 123 TDTVMAY+EGNDSDEIITATAKALVDINEHSNRSGPKRISDVLKVNENLRISSQQTSD 181
 T+VM+AY E +E+TA E+AL+++NEHSNRSGPKRISDVLKVNYE LE RS+QTS+
 Sbjct: 128 TDSVNAYDEILKPEFVIAIGVERALIKLNEHSNRSGPKRISDVLKVNAYEAEARSQQTSN 187
- 50 Query: 182 VTGLPTGFRDLRITTLGHPDQLILAARPAVGKTAPVNLIAQNVGTQKRNPAIFSL 241
 VTGLPTGFRDLD+ITTLGHPDQL+ILAARPAVGKTAPVNLIAQNVGTQK+ VAIFSL
 Sbjct: 188 VTGLPTGFRDLKITTTLGHPDQLVILAARPAVGKTAPVNLIAQNVGTQKRTVAIFSL 247
- 55 Query: 242 GBSLVDRLMAAEGNVDSHSLRTGQLTDQNNVVTIAQALADAPIYIDDTGKIKITEIR 301
 GBSLVDRLMAAEGNVDSHSLRTGQLTDQNNVVTIAQALA+APIYIDDTGPKIKITEIR
 Sbjct: 248 GBSLVDRLMAAEGNVDSHSLRTGQLTDQNNVVTIAQALAKAPIYIDDTGKIKITEIR 307
- Query: 302 ARSRKLSQEVDDGLGLIVIDYQLISGTRPENRQQEVSISRQLKLAKELKVPVIALSQ 361
 ARSRKLSQEV D GLGLIVIDYQLI+GT+PENRQQEVS+ISRQLKLAKELKVPVIALSQ
 Sbjct: 308 ARSRKLSQEVDDGLGLIVIDYQLITGTXPENRQQEVSISRQLKLAKELKVPVIALSQ 367
- Query: 362 LSRGVQRQKRPVLSDIRSGSIBQADIVAFVLYRDTTYRREGEAREIVDMVT+SVIL 421
 LSRGVQRQKRPVLSDIRSGSIBQADIVAFVLYRDTTYR+E +ARE VENT+EVIL

-1161-

Sbjct: 368 LSRGVQRQCKRPVLSDIRESGSIQCADIVAPLYRDDYYRKCDDAEZAVEDNTIEVIL 427

Query: 422 EKRRAGARGTVKLMFQKYNKPSISIAQFEE 451
EKRRAGARGTVKLMFQKYNKPSISIAQFEE

5 Sbjct: 428 EKRRAGARGTVKLMFQKYNKPSISIAQFEE 457

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1039

- 10 A DNA sequence (GBSx1111) was identified in *S.agalactiae* <SEQ ID 3211> which encodes the amino acid sequence <SEQ ID 3212>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4909(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3213> which encodes the amino acid sequence <SEQ ID 3214>. Analysis of this protein sequence reveals the following:

Possible site: 21
>>> Seems to have no N-terminal signal sequence

- 25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3467(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 30 An alignment of the GAS and GBS proteins is shown below.

Identities = 77/90 (85%), Positives = 84/90 (92%)

- 35 Query: 1 MSDAFADVAKMKKIKEDIKSHBQMVLTLENKRKRKKNIGRLIEVYPSLFIVEYKDTA 60
MSDAF DVAKMKKIKEDI++HBQ+VELTLENKRKRKKNIGRLIRVY SLFI+EY D++
Sbjct: 11 MSDAFIDVAKMKKIKEDIHAEGLVELTLENKRKRKKNIGRLIRVYSSLPFIIEYDSS 70

- Query: 61 AVPGAIDNTYVESYTYSDILTEKTLIRYFD 90
FGAIDN+YVESYTYSDILTEKTLIRY D
40 Sbjct: 71 DTFGAIDNSYVESYTYSDILTEKTLIRYLD 100

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1040

- 45 A DNA sequence (GBSx1112) was identified in *S.agalactiae* <SEQ ID 3215> which encodes the amino acid sequence <SEQ ID 3216>. This protein is predicted to be 30S ribosomal protein S4 (rpsD). Analysis of this protein sequence reveals the following:

Possible site: 27
>>> Seems to have no N-terminal signal sequence

- 50 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2937(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-1162-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAC00397 GB:AP008220 ribosomal protein S4 [Bacillus subtilis]
Identities = 138/201 (68%), Positives = 158/201 (77%), Gaps = 1/201 (0%)

Query: 1 MSRYTGPSWKQSRRLGLSLTGTGKELARRNVVPGQHGPNRNSKLSEYGLQLAEKQKLRF 60
M-RYTGPSWK SRRIG+SL+GTGKEL +R Y FG HGP R KISEYGLQL EKQKLR
10 Sbjct: 1 MARYTGPSWKLSRRLGLSLGTGKELKRPYAGPHGPGQRRKLSLEYGLQLAEKQKLRF 60

Query: 61 YGLGEKQFNNLFVQATKAKEGTLGPNFMVLLERRLDNVVYRLGLATTRRQARQFVNHHI 120
YG+ E+QFR LF +A K G G NFM+LL+ RLDMVY+LGLA TRRQARQ VNHGI
15 Sbjct: 61 YGVNERQFRTLPDKAGKLA-GKHSNFMILLDSRLDNVYKLGATRRQARQVNHGI 119

Query: 121 LVDGKRVDIPSYKVTGQVISVREKSMKVPAILLEAVEATLGRPAFVSFDAEKLGGSLTRL 180
LVDG RVDIPSY V PGQ I VREKS + I E+VE P +++FDAEKLGG+ TRI
15 Sbjct: 120 LVDGSRVDIPSYLVKPGQITGVREKSRNLIIKESVEVNNFVPSYLTFDAEKLGGTFTRL 179

Query: 181 PERDEINPEINEALVVEFYNK 201
PER E+ PEINEAL+VEFY++
20 Sbjct: 180 PERSELAPEINEALIVEFYSR 200

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3217> which encodes the amino acid sequence <SEQ ID 3218>. Analysis of this protein sequence reveals the following:

25 Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.2937 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 201/203 (99%), Positives = 201/203 (99%)

35 Query: 1 MSRYTGPSWKQSRRLGLSLTGTGKELARRNVVPGQHGPNRNSKLSEYGLQLAEKQKLRF 60
MSRYTGPSWKQSRRLGLSLTGTGKELARRNVVPGQHGPNRNSKLSEYGLQLAEKQKLRF
Sbjct: 1 MSRYTGPSWKQSRRLGLSLTGTGKELARRNVVPGQHGPNRNSKLSEYGLQLAEKQKLRF 60

40 Query: 61 YGLGEKQFNNLFVQATKAKEGTLGPNFMVLLERRLDNVVYRLGLATTRRQARQFVNHHI 120
YGLGEKQFNNLFVQATK KEGTLGPNFMVLLERRLDNVVYRLGLATTRRQARQFVNHHI
Sbjct: 61 YGLGEKQFNNLFVQATKKEGTLGPNFMVLLERRLDNVVYRLGLATTRRQARQFVNHHI 120

45 Query: 121 LVDGKRVDIPSYKVTGQVISVREKSMKVPAILLEAVEATLGRPAFVSFDAEKLGGSLTRL 180
LVDGKRVDIPSYK FGQVISVREKSMKVPAILLEAVEATLGRPAFVSFDAEKLGGSLTRL
Sbjct: 121 LVDGKRVDIPSYKVDGQVISVREKSMKVPAILLEAVEATLGRPAFVSFDAEKLGGSLTRL 180

Query: 181 PERDEINPEINEALVVEFYNKML 203
PERDEINPEINEALVVEFYNKML
50 Sbjct: 181 PERDEINPEINEALVVEFYNKML 203

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1041

55 A DNA sequence (GBSx1113) was identified in *S.agalactiae* <SEQ ID 3219> which encodes the amino acid sequence <SEQ ID 3220>. Analysis of this protein sequence reveals the following:

Possible site: 29
>>> Seems to have no N-terminal signal sequence

-1163-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4067(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF98302 GB:AF243383 unknown; Orf3 [Lactococcus lactis subsp.

lactis]

Identities = 46/97 (47%), Positives = 69/97 (70%)

Query: 1 MNLNDRLKIEEMEEKYDSFKPRINALVRAIDDFQKHEDYVLRREFYGSDEWFRLSQTE 60
 M+ D I++ME KYD+P P + L+++++ F Y +Y++LR FYGE WF E +
 Sbjct: 1 MDNKKDLIELIQCMENKYDTMPFVLTNLIDSVEKPNSTYNNYIELANFYGSEKWFPEYMEIK 60

Query: 61 NNLLKCGVLSEDQLFDFFIGSHNELVGQFLDMSSQMYRH 97
 +KCGVL+EDQLFD I +HNEL+G LD++S+MY++
 Sbjct: 61 LPVKCGVLTEQLFDWISDHNELLGVLDDLTSKMYKN 97

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3221> which encodes the amino acid sequence <SEQ ID 3222>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3465(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 48/98 (48%), Positives = 74/98 (74%)

Query: 1 MNLNDRLKIEEMEEKYDSFKPRINALVRAIDDFQKHEDYVLRREFYGSDEWFRLSQTE 60
 M D+L +E+ME+ Y++F P++ L+EA+D F++HYE+Y LR FY S++WFRLL+ Q
 Sbjct: 1 MTKQDQLIVERMEQTYEAFSEKLANLIRALDAFKEHYEETATLRNFYSSDEWFRLLANQEW 60

Query: 61 NNLLKCGVLSEDQLFDFFIGSHNELVGQFLDMSSQMYRH 98
 +++ CGVLSED LFD IG+HN+L+ LD++ MY+H+
 Sbjct: 61 DDIPCGVLSEDLDFWIGDHNGLLADLIDLAPIMYKIM 98

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1042

A DNA sequence (GBSx1114) was identified in *S.agalactiae* <SEQ ID 3223> which encodes the amino acid sequence <SEQ ID 3224>. Analysis of this protein sequence reveals the following:

Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0965(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:EAB04438 GB:AP001509 transcriptional regulator (TetR/AcrR

family) [Bacillus halodurans]

Identities = 47/181 (25%), Positives = 95/181 (51%), Gaps = 16/181 (8%)

Query: 4 DTRREKTKRAIEAMITLLKQSPDEISTINI/TKTAGISRSSFYTHYDKYBMIDCYQGS 63

-1164-

D R++ T+ ++ +++ L++++ I+ + A I+RS+FY+HY D Y+++ Q +
 Sbjct: 6 DRRKKYTRMLLKESLMKLAQEKPLSNITIKICDLADINRSTFYSHYTDLYDLAYQIEDS 65
 Query: 64 LFNKV-EYIFDRNQFKKEDAL-----LEIPQFLDRESLFAALLTQNGTKIKIQTYYIANKQ 117
 + + E+ + N K E+AL L ++ +RES L ++ G Q K
 Sbjct: 66 LIXDLSEALSSNYTKDEEALQMTENLLVYIANWRSC-QTLFSEYGDPSFQ----KKV 119
 Query: 118 LMLSKELPVNFP--DATKSDINRLYSYVLSHAIFGVYQMWITRGKESPOQITQVLLSL 175
 +ML+ + + P TK DI+ Y S+Y+ + + Q W+ G K+SP+++ +++ L
 Sbjct: 120 MMLAHDRVIKTPLVGKHFKPDISE-YVSLYIVNGSIHIVQSWLKNGLQSPKEMAEILIKL 179

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3225> which encodes the amino acid sequence <SEQ ID 3226>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have an uncleavable N-term signal seq
 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAB04438 (GB:AP001509 transcriptional regulator (TetR/AcrR family) [Bacillus halodurans])
 Identities = 47/180 (26%), Positives = 88/180 (48%), Gaps = 18/180 (10%)
 Query: 4 RKENTQAILKAWMLLKTESFDDITTVKLSKRAGISRSSFYTHYKDYEMIDYQITFF 63
 RK+ T+ +++++ L+++ +IT ++ A I+RS+FY+HY D Y+++ +
 Sbjct: 8 RKKYTRMLLKESLMKLAQEKPLSNITIKICDLADINRSTFYSHYTDLYDLAYQIEDS 67
 Query: 64 KLEYIFEKKYQNKQAPLEVFPEPL-----QRELISLISANGTKETQAFIINKVRLI- 117
 L K++ I++ E L + +L S G Q KV +L
 Sbjct: 68 KDLSEALSSNYTKDEEALQMTENLLVYIANWRSCQTLFSEYGDPSFQ----KKVMLA 123
 Query: 118 -----ITTDLDQKPFSTEELQTEKEYQSYLAHAPFGVCQSHIANGKESPOEMTQVLM 173
 I T L K + ++S EY S+Y+ + + QW+ G K+SP+EM + ++K+
 Sbjct: 124 HDRVIKTPLVGKHFKPDISE----EYVSLYIVNGSIHIVQSWLKNGLQSPKEMAEILIKL 179

An alignment of the GAS and GBS proteins is shown below.

Identities = 100/179 (55%), Positives = 134/179 (73%), Gaps = 2/179 (1%)
 Query: 1 MVDTRREKTKRAIFAMITLNDQSPDEISTINLTKTAGISRSSFYTHYKDYEMIDY 60
 MVN R+E TK-AI A+ LK +SPD+I+T+ L+K AGISRSSFYTHYKDYEMID Y
 Sbjct: 1 MVN--RKENTQAILKAWMLLKTESFDDITTVKLSKRAGISRSSFYTHYKDYEMIDY 58
 Query: 61 QQSLFNKV-EYIFDRNQFKKEDALLEIPQFLDRESLFAALLTQNGTKIKIQTYYIANKQ 120
 QQ+ F+K-EYIP++ KE A LE+ F+FL RE L ++LL+ NTKETIQ +I+NK+L++
 Sbjct: 59 QQTFFHKKLEYIFEKKYQNKQAPLEVFPEPLQRELISLISANGTKETQAFIINKVRLI 118
 Query: 121 SKELPVNFPDATKSDINRLYSYVLSHAIFGVYQMWITRGKESPOQITQVLLSLPQT 179
 + +L S + Y S+Y+HA FGV Q WI +GKESPO+TQ +L +L T
 Sbjct: 119 TTDLDQKPFSTEELQTEKEYQSYLAHAPFGVCQSWIAKGGKESPOEMTQVFLMLTST 177

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1043

A DNA sequence (GBSx1115) was identified in *S.agalactiae* <SEQ ID 3227> which encodes the amino acid sequence <SEQ ID 3228>. Analysis of this protein sequence reveals the following:

Possible site: 58

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.35	Transmembrane	790 - 806 (787 - 808)
INTEGRAL	Likelihood = -7.32	Transmembrane	707 - 723 (703 - 725)
INTEGRAL	Likelihood = -7.11	Transmembrane	637 - 653 (630 - 659)
INTEGRAL	Likelihood = -6.32	Transmembrane	678 - 694 (672 - 698)
INTEGRAL	Likelihood = -1.44	Transmembrane	55 - 71 (55 - 73)
INTEGRAL	Likelihood = -0.22	Transmembrane	732 - 748 (730 - 748)

----- Final Results -----

bacterial membrane ---	Certainty=0.5140 (Affirmative) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm ---	Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10287> which encodes amino acid sequence <SEQ ID 10288> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12856 (GB:Z99109 alternate gene name: yixE-similar to phage infection protein [Bacillus subtilis])

Identities = 227/783 (28%), Positives = 387/783 (48%), Gaps = 60/783 (7%)

Query: 45 KAIKSPKLMITMAGVALIPTLYNVIFLSMMWDPYGNINLNPVAVVNDQSAKLNKGTIS 104
K I + S K L I + + + P + Y + + F L + W D P Y G L P V V V N Q D K A G + +

Sbjct: 9 KDIVTSKKLLIPILIAILFVPLIYSGVFLKAYWDPYGTVDQLPVTVVNDQSGATYEGKQLQ 68

Query: 105 IGRIMEDNLSKNDSDIDPHFTT-AQRARKELEKGHYYMVITFPKDLRSKATITLMTKEPERL 163
I G D + L N + + D + H F + + K + L Y Y + V + P + D S + A + T + + + P + L +

Sbjct: 69 IGRDLVKELKDNNNFDFWHFSNDLDQSLKDLNCKYLYLVVEIPEDFSKNASTVLDKPNPKL 128

Query: 164 NITYKTTKGRSFSVASKMSETAANKLKDEVAESITGTYTESVFQNMKSMKGTINKAADGSG 223
+ + Y T G + + V + + E A + K L K V + + T Y T + F N + G + + A + G + +

Sbjct: 129 DLICHTNAGNSYVGATIGEKALDKLQASVSEKVEYQYTKVIFDNFKDIAKGLSDASSGK 188

Query: 224 ELLNGSNKLQDGSQTLTSLNLDVLASSQTFSGGANKLNSGINLYTDGVGTLSNGLTSLD 283
+ + + G + + G S L N L L S + T S + L G T G + + L + L D

Sbjct: 189 KIDGTFKDAKNGSAQLKENIAKLKCESTATISDKTQALADGAAQVTSIGIQSLDSLQKPD 248

Query: 284 GVTAYTTGVHKLSEGSQKLDKSKQALV-----EGSEKLTDLQLQLSQATQLKPEQRT 336
+ L + G S + L K L + + G + L T + G L Q L + Q E +

Sbjct: 249 SBNQIYDKSSQLAAGSGELTSKQVIELLAGLQNVQKTFPHLTNGLDQLNSKVQBGSEFAK 308

Query: 337 LQNLSDG--LKNLNQIITNLQSTATTDSBTNSKLFNPLSTIESSTKALMPTAARDKQKM 394
+ + + L L + N L + + T + + L + F + + + + + A N + + +

Sbjct: 309 AEKIINALDITKLETAANNLEKSEATAMKEFKQLDTFENSLKRDQAFKN--VINSNDF 366

Query: 395 TAVQST----SAFKSLTPEQQSQITSAVTGFTPSAE-TIAANISSNIENMKTVLSEASS 449
T A Q + S K L + + P T + A I S + + E + K + + +

Sbjct: 367 TAEQSQQLNSVEKKLPQVDAPDFDQILSQLEPTADQLPDIAITKSLIEDVQAQVQKAM 426

Query: 450 APEN---NGSQNLQTLGTTANNVLKATSDLDKIQKLPATKQLYQSSQTLTKGITYDT 505
+ N G + + Q D I + L + + Y G S Q L T G T T

Sbjct: 427 PEATSKLYNGAKTIQ-----DAIDRLTEGADKIYNGSKQLIDGQTKLT 469

Query: 506 NAVQQLRKGAIVLDSKSNQLISGTQKASQCAQTLDSKSDQLRDGAGSLAGSDRIADGNS 565
+ G + K + S Q L + G S Q + G + L G S + + G S +

Sbjct: 470 AGIGSEYNKQFAKAKAGSEQLATG-----SSQVSGGLFKLLDGSQKQVSGS 515

Query: 566 KLAGGGHQLDTGLTSLSGVSQLSSSLGKAGDQLSMVSVKNKDNANAVSPVTIKHEDYDS 625
K L A G L G L + L G + L S S L A D Q + + + P V K + S

Sbjct: 516 KLAGDSASLDITGLKGLDGTGETLSSKLKDAADQTDGIDADDQTYGMFADPVTIKDAHS 575

Query: 626 VDTNGSVTAAPYMSVALMVVALSANVIFAKALSGKEPANRFSWAKNK--LLINGFIATL 682
V G G + P Y + S + L V + V + F + P N F W + K + + + G I + L

Sbjct: 576 VPNYGTGLTPYILMGLVYVGIMITVVFPLKASGRPRNGFENFSPKFNVMKLVGIQSL 635

Query: 683 -AATILFVAFQFGLKPDYPORTYFILLTAWTLMALVATLVGMWDRNYSGLSILLILFQ 741

-1166-

AT+L IGL+ + + Y ++T+ +A++ L G P+++IL+ Q
 5 Query: 636 IVATVLLIG---IGLEVESTWRFVFTIITSIAFLAIQFLATTMGNPGRFIAVILVILQ 692
 742 LQSSAGTTPPELSPKPFQFIQPIPLMPTYSVSGLETTISLTGDVNHQRMVIFLVSSMIL 801
 5 LQ+S GT+P+EL P F+Q I LPMPTYS++G R IS GD +W+M + + ++++
 693 LQSSGSGTTPPELLEPNFYQVHGALEPMTYSINGFRAVIS-NGDFGYMAGVGLIGLIALVM 751
 Query: 802 ALL 804
 L
 10 Sbjct: 752 DAL 754

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2017> which encodes the amino acid sequence <SEQ ID 2018>. Analysis of this protein sequence reveals the following:

Possible site: 26
 15 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.29 Transmembrane 735 - 751 (729 - 754)
 INTEGRAL Likelihood = -5.79 Transmembrane 582 - 598 (580 - 601)
 INTEGRAL Likelihood = -3.66 Transmembrane 652 - 668 (650 - 669)
 INTEGRAL Likelihood = -2.97 Transmembrane 14 - 30 (14 - 34)
 20 INTEGRAL Likelihood = -2.66 Transmembrane 623 - 639 (622 - 641)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4715 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 360/779 (46%), Positives = 508/779 (66%), Gaps = 32/779 (4%)
 30 Query: 40 MLDELKAIKSPKLNITMAGVALIPTLVNIVFLSSWMDPFGNTGNLVPVAVVQOKAKLN 99
 1 ML+ELK +IK+PKL IIM GVAL+P LTN+ FL SMDPFG +LP+AVVN DK AKR 60
 Sbjct: 1 MLEELKTLIKNPKLIMTIGVALVPAVNLISFLGSMDFYGRVNDLPIAVNHHOKPFA 60
 Query: 100 GKTISIGKMBEHLNSKIDSLDFHTTAKRAEKLEKGHYTMVITFPKDLRKKATTMTTEK 159
 35 K++IG DM D +SK+ L+HF +AK+AA+ L+G YMVIT +P+LS+AA TL+ +
 61 DKSLITGDMVDKMSKSKDLHYHFVSARKQAGLEKEGDTYTNVITLPEDLSQRAATLTAPE 120
 Query: 160 PERIMITYKTKGRSFVASKASSETAANKLKDEVAESITOTYTESVFQMGSMKGTINKAA 219
 40 P+L I Y+T+K3 VA+KM ETA KKK+ V++IT TYT +VP +M +++G+ +A
 121 PQKLITIRYQTSKGHWVAIKGSETAMAKLKSVSQNTIKTYTSVAFSSMTDLQSGLEKAS 180
 Query: 220 DGSQRLANGSKNLQDGSQTLTSMILVLAASSQTFPGGANKLNGINLYTDGVTGLNKGL 279
 45 DSQ L +G+ Q GSQTL+NL L +SQ F +G +L S+ YTDGV +NGL
 181 AGSQALASCAKTAQGSQTLSTNLAAITGASQFQQGTGRLTGSLTTYTDGVNQGVLG 240
 Query: 280 TLDSDVYATYTVGHVLESGSKQLDDKQSALVEGSEKLTGQLQSLQATLKPSEKRTLQ 339
 50 TLS + Y GV +LG+G+ +L+ GL QL+QAT L E+ + +Q+
 241 TLDSDIPNYLNGVSRLSQGSQNLN-----GLSQLTQATLTSDKAKGIQS 286
 Query: 340 LSGDKMLNQITNLQSTATTDSYTN--SKLPFLSTIESSTKALMFAADKQKQNTA 396
 60 L GL LNQ I L + +T N +NLN I + K++ A +++++A
 287 LIVGLPVLNGGIQQLNLTSLQPPNANADKLNGSLGATAQAQKQVIARETAQNEELISA 346
 Query: 397 VQSTSAFKSLTPEQQSITGAVTGTPTSAETIAAN-ISSNIENMCTVLSSESSAPSRNG 455
 55 +Q+TS ++SLT EOO ++ +A++ + S AA I S+++ +TL S S
 347 LQTSVYQSILTAQQGELAAALSQSDKSTVSAQITLSSVQTLSTLSQSLSQSDQSQKL 406
 Query: 456 SQNLGTLSTANNVLKASIDLDKIKLPATKOLYQSSQTLTKGIDTYTNV-----GQL 521
 60 Q + + + AN Q LP A+ L + S L K V
 407 BQLKEAVAQIANQ-----SNQALPQASSALTELSTGLAKVNGSLNQVLPQSNQL 456
 Query: 512 RKGAVITLDSKSNQILSTQKASQGAITLDSKSDQLRDGAGQLASGSDRIADGSKNLKAGG 571
 65 G L+ + + SG K+S+GA L SKS +L DG+ OL+ G+ ++ADGS++L+ GG
 457 TTTLAQNLNRYNTAIGSGVILKEGANAALSSKSGELLDGSHGLSEBQATKLADSSQSLQSG 516

-1167-

- Query: 572 HQLTDLGLTELSGVSGLSSSLGKAGDOLSMVSVNKNANAVSPVTIKHEDYDSVDTMGV 631
 HQLT GLTELS G4S L+ SL KA QLS+VSV NA AV+ P+ + +D D V TNG+
 Sbjct: 517 HQLTSGLTELTGLSTLNGSLAKASQQLSLVSVTDKNKAVKAPLVINEKDQGVINGI 576
- 5 Query: 632 GNAPYMI+V+IMVVALS NVIFA +LEG+ +++ WAK K +INGFI+T+ + +L+ A+
 GNAPYMI+V+IMVVALS NVIFA +LEG+ +++ WAK K +INGFI+T+ + +L+ A+
 Sbjct: 577 GNAPYMI+V+IMVVALS NVIFA +LEG+ +++ WAK K +INGFI+T+ + +L+ A+ 636
- 10 Query: 692 QFGLKPEYEGKTYPIILLTANTLMAVLTALVGNQNRYSFLSLILLPLQSSAGTYPI 751
 Q +G + Y +T I+L+ WT MALVTPALVGNQ+RYGSP SL++LQ Q+GSS G+YPI
 Sbjct: 637 QLGLFEARYGMEITGLPTMSGWTFMALVTPALVGNQNRYSFASLVMLLQVSGSGSYPI 696
- Query: 752 ELSPKFPQTIQPLPMYTSVSGLRITISLTGDVNHQKRWLVIFLVSSMIALLIYRKQR 810
 ELS FFO + PFLPMYTSVSGLR+TISL+G + +L FL++ M+LALLIYR ++
 15 Sbjct: 697 ELSGAFFQKLHPFLPMYTVVSGLRQTISLSCHIGVEVKVLTGFLFAMVIALLIYRPPK 755

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1044

- 20 A DNA sequence (GBSx1116) was identified in *S.agalactiae* <SEQ ID 3229> which encodes the amino acid sequence <SEQ ID 3230>. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence

- 25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2664 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1045

- 35 A DNA sequence (GBSx1117) was identified in *S.agalactiae* <SEQ ID 3231> which encodes the amino acid sequence <SEQ ID 3232>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.45 Transmembrane 48 - 64 (45 - 69)
 40 INTEGRAL Likelihood = -1.49 Transmembrane 71 - 87 (71 - 87)

- Final Results -----
 bacterial membrane --- Certainty=0.4779 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 45 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9441> which encodes amino acid sequence <SEQ ID 9442> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

- 50 >GP:AAA25222 GB:M87483 ORF 1 [Lactococcus lactis]
 Identities = 50/88 (56%), Positives = 66/88 (74%), Gaps = 1/88 (1%)
 Query: 2 TGRIFMSKEELSYLPIKLFKNQGVINGLIGLFLYGLYISQNG-EIVAVFLINVLVA 60

T ++F+M KEEL V LFNKQ+YNGLIGL L+Y ++ S Q EIV + LI ++LVA
 Sbjct: 32 TSRVFNMGKEELERSSVQLTFKNQGIYNGLIGLGLIYAFFSSACLEIVRELLLIYIILVA 91
 Query: 61 IYGALTVDKIKILLKQSGLPILALITPLF 88
 +YG+LT ++KKI+L QGGL ILAL+ F
 Sbjct: 92 LYGSLTNNKIIITOGGLALIALISSSF 119

No corresponding DNA sequence was identified in *S. pyogenes*.

A related GBS gene <SEQ ID 8719> and protein <SEQ ID 8720> were also identified. Analysis of this protein sequence reveals the following:

```

LipoP: Possible site: -1      Crend: 8
MoG: Discrim Score:         4.19
GVH: Signal Score (-7.5): -3.99
    Possible site: 38
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 3 value: -9.45 threshold: 0.0
INTEGRAL      Likelihood = -9.45   Transmembrane    87 - 103 ( 84 - 108)
INTEGRAL      Likelihood = -1.49   Transmembrane   110 - 126 ( 110 - 126)
INTEGRAL      Likelihood = -0.37   Transmembrane   13 - 29 ( 13 - 29)
PERIPHERAL    Likelihood = 0.47
                                     65
modified ALOM score: 2.39

*** Reasoning Step: 3

----- Final Results -----
bacterial membrane --- Certainty=0.4779(Affirmative) < success
bacterial outside --- Certainty=0.0000(Not Clear) < success
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < success

```

The protein has homology with the following sequences in the databases:

[illegible]

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1046

A DNA sequence (GBSx1118) was identified in *S. galactiae* <SEQ ID 3233> which encodes the amino acid sequence <SEQ ID 3234>. Analysis of this protein sequence reveals the following:

Possible site: 41

-1169-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 5 bacterial cytoplasm --- Certainty=0.3140(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10285> which encodes amino acid sequence <SEQ ID 10286> was also identified.

- 10 The protein has homology with the following sequences in the GENPEPT database.

>GP:CB12447 GB:Z99107 similar to arylesterase [Bacillus subtilis]
 Identities = 37/91 (40%), Positives = 56/91 (60%)

- 15 Query: 13 KDGSDIYRVRVGGQPIVFLKNGSLSSRYFDKQLIAYPSKYTVIWMDSRGHGKSHAKLNT 72
 +D + +YY G G PI+F+HG +S ++F KQ + S YQ I +D RGHG+S L+
 Sbjct: 7 EDQTRLFYETHGSGTPIFLFHGVLMGSGQFHHKQFVLSANTYQICRLDLRGHSGSDKVLHG 66

- Query: 73 ISFRQAVDLKDIWLHLEIDKVIIVGHSDGA 103
 + Q A D+++ L +E+D V+L G S GA
 20 Sbjct: 67 HTISQYARDIREFLAMELDHVILVAGWSGA 97

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1047

A DNA sequence (GBSx1119) was identified in *S.galactiae* <SEQ ID 3235> which encodes the amino acid sequence <SEQ ID 3236>. This protein is predicted to be an integral membrane protein. Analysis of this protein sequence reveals the following:

Possible site: 58

- 30 >>> Seems to have an uncleavable N-term signal seq
- | | | | |
|-------------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -12.90 | Transmembrane | 14 - 30 (9 - 41) |
| INTEGRAL | Likelihood = -9.71 | Transmembrane | 451 - 467 (447 - 472) |
| INTEGRAL | Likelihood = -9.18 | Transmembrane | 234 - 250 (229 - 257) |
| INTEGRAL | Likelihood = -8.07 | Transmembrane | 56 - 72 (46 - 77) |
| 35 INTEGRAL | Likelihood = -8.01 | Transmembrane | 490 - 506 (484 - 512) |
| INTEGRAL | Likelihood = -5.84 | Transmembrane | 414 - 430 (412 - 436) |
| INTEGRAL | Likelihood = -4.99 | Transmembrane | 136 - 152 (135 - 159) |
| INTEGRAL | Likelihood = -4.14 | Transmembrane | 213 - 229 (211 - 232) |
| 40 INTEGRAL | Likelihood = -4.14 | Transmembrane | 365 - 381 (364 - 382) |
| INTEGRAL | Likelihood = -2.66 | Transmembrane | 393 - 409 (391 - 412) |
| INTEGRAL | Likelihood = -1.06 | Transmembrane | 168 - 184 (167 - 184) |
| INTEGRAL | Likelihood = -0.64 | Transmembrane | 275 - 291 (275 - 291) |
| INTEGRAL | Likelihood = -0.32 | Transmembrane | 328 - 344 (328 - 345) |
| 45 INTEGRAL | Likelihood = -0.27 | Transmembrane | 821 - 837 (821 - 837) |

----- Final Results -----

- 50 bacterial membrane --- Certainty=0.6158(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10283> which encodes amino acid sequence <SEQ ID 10284> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

- 55 >GP:BA024464 GB:D85082 YfiX [Bacillus subtilis]
 Identities = 190/596 (31%), Positives = 324/596 (53%), Gaps = 31/596 (5%)

Query: 246 IVSLIPGGGLGSPFLVLTGFAEGLPKSTVAVMLLYRLAYYIIPFFAGIYFFIHYLSGQ 305
 ++SL+PGG GSF+L+ G G +E +V ++LYRLAY IPF G++F L
 Sbjct: 1 MISLAVPGGFSFDLLFLGMEQLGYHQAIVTSIVLYRLAYSFIPIFLGLFPAAGLDITEN 60

Query: 306 INQRYENVPK-----ELVSTVLQTMVSHLRILG---AFLIFSTAFFENITYIMMLQKLG 357
 +R E P+ E ++L + L+RIL + ++F + + + +L
 Sbjct: 61 TMKRLTNPRIAPAEITNNVLLVQRAVLRLIQQSLSLIVFVAGLIVLASVSLPIDRIT 120

Query: 358 LDP-LQEQMLWQFPGLLIGVCFILARTID--QKVNAPPIAIWITI/LFYINLGHISW 414
 + P + L F GL L ILL I+ ++K ++AI + + L ++
 Sbjct: 121 VIPHPRPALLNLNGLSSLLAILLILPELYKRTKRSYTMALTAIVGGPFVSLKGLNI 180

Query: 415 RLSFWFILLLLGLLVKPTLYKKQPIYSWEERIKDGLIIVSLMGVLFY----YAGLFPPI 470
 F +++ L++K ++Q Y+ + I V+L V + IAG ++
 Sbjct: 181 SAIFVLPMIILVLLKQFVRBOASYTLGOLI----FAVALFTVALFNYNLIAAGFWDR 236

Query: 471 RAHTVGGSIERLHYIAMEPIALATL----ILTLVYLCLVKILQKSKQIGDVNVDRYK 526
 + + +++ + I AT+ I+L+L + ++ IG+ +R
 Sbjct: 237 MKKV---LRHEYFHSHTSHITHATIMAILIIVPLFIPTVVYHKRTKPIGRKADPERIA 292

Query: 527 KLIGAYGGSSDSGLAFINDKRLYWKNGEDCVAFQFVIVNNKCLINGEPAGDDITYIREA 586
 L GG++ S L FL DKR Y + +G + F + + ++G+P+G
 Sbjct: 293 APLINEKGNALSHLGLGDKRFP--PSSDGNALLPLGKIA--RLRVVLGDPGSGQSFPLV 349

Query: 587 IRSFIIDADKLIDYLVFYSIGQKLTLLHEYGDPFMKVGEDALVNLEFTLKGKNYKPKFR 646
 +E F++A + + ++FY I ++ L H++G++F K+G+A V+L TPTL G K R
 Sbjct: 350 LEEFLNEAHQKGSVLFPQIEREDMALYHDFGNFPLGEEAYVDINTPTLTKGKAGLR 409

Query: 647 NALNRVSKDGFYEVVQSPHSOELINSLEEISNTWLEGRPEKGPSLGYNKDYPCQAPIA 706
 NR E++ + F V P S L L++IS+ WL + EKGPSLG+P+ Y Q+ADPIA
 Sbjct: 410 AIDNRFSREYTFPHDHPFSDAFLEELKQISDWMGLSKCKEKGPSLGFPDPSYLGKLR 469

Query: 707 LVQNAEHEVVAEAMIMPYEKSIIISIDLMRHDQKIPANGVMDFLSLFSYYQEKGYHYF 766
 +QNAE +VAEAM+NP Y++ IS+DLMR+ + PNG+MD LF+ +F + +E+G F
 Sbjct: 470 YNKAGBEIVAFAMNMPYQGEISVDLMR--RGDAPNGINDALFIRMFMAKEBGTCSF 528

Query: 767 DLQMAPLSGVGRVETSPAKERMAYLVYHFGSHFYSPNGLHKYKKFTPLMSERYIS 822
 ++GMAPL+ VG TSP ER A +++ + YSP+GL +K+K+ P W +Y++
 Sbjct: 529 NMGMAPLANVGTAPTSFWSERFAAVI PNVVRYMYSFSLGRAPFKYKPEWRGKYLA 584

No corresponding DNA sequence was identified in *S. pyogenes*.

A related GBS gene <SEQ ID 8721> and protein <SEQ ID 8722> were also identified. Analysis of this protein sequence reveals the following:

45 Lipop: Possible site: -1 Crend: 9
 MG: Discrim Score: 9.22
 GVH: Signal Score (-7.5): -7.66
 Possible site: 58
 >>> Seems to have an uncleavable N-term signal seq

50 ALOM program count: 14 value: -12.90 threshold: 0.0

INTEGRAL	Likelihood = -12.90	Transmembrane	14 - 30 (9 - 41)
INTEGRAL	Likelihood = -9.71	Transmembrane	451 - 467 (447 - 472)
INTEGRAL	Likelihood = -9.18	Transmembrane	234 - 250 (229 - 257)
INTEGRAL	Likelihood = -8.07	Transmembrane	56 - 72 (46 - 77)
INTEGRAL	Likelihood = -8.01	Transmembrane	490 - 506 (484 - 512)
INTEGRAL	Likelihood = -5.84	Transmembrane	414 - 430 (412 - 436)
INTEGRAL	Likelihood = -4.99	Transmembrane	136 - 152 (135 - 159)
INTEGRAL	Likelihood = -4.14	Transmembrane	213 - 229 (211 - 232)
INTEGRAL	Likelihood = -4.14	Transmembrane	365 - 381 (364 - 382)
INTEGRAL	Likelihood = -2.66	Transmembrane	393 - 409 (391 - 412)
INTEGRAL	Likelihood = -1.06	Transmembrane	168 - 184 (167 - 184)
INTEGRAL	Likelihood = -0.64	Transmembrane	275 - 291 (275 - 291)
INTEGRAL	Likelihood = -0.32	Transmembrane	328 - 344 (328 - 345)
INTEGRAL	Likelihood = -0.27	Transmembrane	821 - 837 (821 - 837)
PERIPHERAL	Likelihood = 1.06		558

65

-1172-

AVIPNNVRYMYSFSGILRAFKKCYKPEWRGKYLAYRQRSLSVIMFLVTRLIIGKSKKDSV
 610 620 630 640 650

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1048

A DNA sequence (GBSx1120) was identified in *S. agalactiae* <SEQ ID 3237> which encodes the amino acid sequence <SEQ ID 3238>. This protein is predicted to be choline transporter. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.24	Transmembrane	28 - 44 (22 - 47)
INTEGRAL	Likelihood = -8.81	Transmembrane	178 - 194 (176 - 204)
INTEGRAL	Likelihood = -7.22	Transmembrane	81 - 97 (63 - 105)
INTEGRAL	Likelihood = -3.50	Transmembrane	209 - 225 (206 - 226)
INTEGRAL	Likelihood = -3.13	Transmembrane	64 - 80 (63 - 80)
INTEGRAL	Likelihood = -2.44	Transmembrane	156 - 172 (153 - 172)
INTEGRAL	Likelihood = -0.64	Transmembrane	137 - 153 (137 - 153)

----- Final Results -----
 bacterial membrane --- Certainty=0.5097 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AD45530 GB:AF162656 choline transporter (Streptococcus pneumoniae)
 Identities = 326/505 (64%), Positives = 409/505 (80%), Gaps = 1/505 (0%)

Query: 1 MTLLITTFQERFGDWQTSLEHQLSLTLILATLIAIPGLIIISHYKKISHVVLQITGI 60
 MT LI TFQ+RF DW +L +RLQLSLTL+LA L+AIPL + + ++K++ VLQI GI
 Sbjct: 1 MTLIATFQDRFSDWLTALSGHQLSLTLILLALLIAPLAVFLRYKHLADNVLAGI 60

Query: 61 FQTIPSLALLGLFIPPMGIGTVPVAVALLIYALFFILQNTVTVLMQIDANLIEAATPFGM 120
 FQTIPSLALLGLFIP MGIGT+PA+ AL+IYA+FFILQNT+T L ID NL EA APGM
 Sbjct: 61 FQTIPSLALLGLFIPMGIGTLPALTALVIYATFFILQNTITGLKGDIPNLQAGIAPGM 120

Query: 121 TRWERLKKFELALSMFVLIISGIRTSVMIIGTATLASLIAGGLGSFILLGIDRNNPSLI 180
 TRWERLKKFE+ L+MPVI+SGIRTA+V+IIGTATLA+LIAGGLGSFILLGIDRNN SLI
 Sbjct: 121 TRWERLKKFRIPLMPVIMSGIRTAUULIIGTATLAALIAGGLGSFILLGIDRNNASLI 180

Query: 181 LIGAISSAVLAIIFSGILGLEKARLRTIAVSGILLIAGLGLSYPKWMPTGTNTATTIVA 240
 LIG+SSAVLAI F+ L+ ++EKA+LRTI L+ LGLSY+P + + +A
 Sbjct: 181 LIGALSSAVLAIAPNLLKVMKEKALRTIPSGFALVALLIGLSYSPALLVQKEKENVLIA 240

Query: 241 GKLGTRPDILINMYKELIEDQTDIKVKLKNFGKTFLYQALKSGDIDLYPEFTGTTSS 300
 GK+G RP+IL NMYK LIE+ T + +KPNFGKT+FLY+ALK GDID+YPRFTGT+T S
 Sbjct: 241 GKIGPEPILINMYKLLIEENTSMATVKNFGKTSFLYALKKKDIDYPEFTGTVTS 300

Query: 301 LLKNPKVSNPNKQVYNIAKNGILKQOKLSLSPMAYQNTYAVAVKKDYAEANQKNISD 360
 LL+ PKVS+ P+QVY +A+GI QKD L+ L PM+YQNTYAVAV K A+ LK ISD
 Sbjct: 301 LLQSPKVSHEPEQVQVARDGIAQDHLAYLKKMSYQNTYAVAVPKYQAGIQSGDITQIDAYS 360

Query: 361 LKKLD-KLKGFTLEFKDREDSIGLQHYGLNLDISTLEPALRYQAINSKDVNIDAYS 419
 LKK++ +LKGFTLEF DREDG+ GLQ YGLN+++T+EPALRYQAI S D+ I DAYS
 Sbjct: 361 LKKVEGLKKGFTLEFNDRDENGKGLQSMYGLNLVATTEPALRYQAGIQSGDITQIDAYS 420

Query: 420 TDSBLIQYQLGILKDKHLFFPPYQGAFLRQDTIKKYPQVKKALNKGHITKEKQBN 479
 TD+EL +Y LQ+L+DDK LFFPPYQGAFL+++ +KK+P++++ LN LAG ITE +M ++N
 Sbjct: 421 TDSBLERYDLQVLRDDKQLFFPPYQGAFLMKRALKKHPLELRVNLTAGKITSSQSQLN 480

Query: 480 YQVAVKHKSAAATVAKQYLKAHHI 504

-1173-

YQV V+ KSA VAK++L+ ++K
 Sbjct: 481 YQGVGEGSAKQVAKFLEQEGLLK 505

There is also homology to SEQ ID 636.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1049

A DNA sequence (GBSx1121) was identified in *S. agalactiae* <SEQ ID 3239> which encodes the amino acid sequence <SEQ ID 3240>. This protein is predicted to be choline transporter (opuBA). Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2345 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AA45529 GB:AF162655 choline transporter [Streptococcus pneumoniae]
 Identities = 139/236 (58%), Positives = 178/236 (74%)

Query: 1 MISFENVSKSYGDIITIDNISCHIQGEFFVLVGASGSGKTTILKMINRLIEPSQGAITL 60
 MI ++WV+ Y + ++ ++ I+ GRF VLVG SSGSKTT+LKMINRL+EP+ G I +

25 Sbjct: 1 MIEYGVNVALRYTEKDVLRDVLNQIEDGEFVVLVGFSGSGKTTMLKMINRLLEPTDGNITM 60

Query: 61 DGENITSLDLRLQLLETGYVLLQQAIALFPNLTVGSNITELIPENKGNKSGDKKKAASDLLDK 120
 DG+ I D R+RL TGYVLQ IALFPNLTV ENI LIPENKGNK + K +LL K

30 Sbjct: 61 DGRKIKDYDERELRSTGYVLLQQAIALFPNLTVAENIALIPENKGNKSGEITKTEELAK 120

Query: 121 VGLPAKDYFNRYPHLSGGEQQRIIGILRAIVAKPKVLLMDPEPFSALDPISRRLQDITKQ 180
 VGLP +Y +R P ELSGGEQQR+GI+RA++ +PK+ LMDEPFSALD ISR+QLQ +TK+

35 Sbjct: 121 VGLPVAEYGHRLPSELGGEQQRVOIVRAMIGQPKIFLMDEPFSALDAISRQLQVLTK 180

Query: 181 LQSELGITLVFVTHMKEAMRLADRICVIEKGIQVQLRPEIIONNPSDQFVRTLF 236
 L E G+T +FVTHD EA++LADRI V+++G+I Q+ PE I P+ FV LF

Sbjct: 181 LHKEFGMTTTFVTHDTEALKLADRIAVLQDGEIRQVANPETILKAPATDFVADLF 236

There is also homology to SEQ ID 644.

- 40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1050

A DNA sequence (GBSx1122) was identified in *S. agalactiae* <SEQ ID 3241> which encodes the amino acid sequence <SEQ ID 3242>. This protein is predicted to be two-component response regulator. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -5.52 Transmembrane 49 - 65 (46 - 66)

----- Final Results -----

bacterial membrane --- Certainty=0.3208 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GF:BA06434 GB:AP001516 two-component response regulator [Bacillus halodurans]
Identities = 101/305 (33%), Positives = 152/305 (49%), Gaps = 31/305 (10%)

```

Query: 1  MKFYIIDDPTITMILQDIEE-EDFNVTVVVRVNVSSKAYNELLIADVDIVLIDLMF 59
      M F+I DDD T+ IL IIR E V + S L I VDI+LIDLMP
Sbjct: 1  MNFPTIDDDVTVRSLAQIIEBQLGQVVGSEADGSHLDGKRLNITKQVDILLIDLMENC 60

Query: 60  DGVTLVQKIYKQRSLKFMISQVKINDLRQRAYKAGIEFFINKPINIIEVKSVMKRVTD 119
      DG+ +QKI K K IMISQ++ +L EAY GIE +I KPIN IEV SV+++V +
Sbjct: 61  DGLRAIQKI -KPEFGKLIIMSQTESKRLSEAYLLGIRHYIMKPINKIEVLSVIRKVIN 119

Query: 120  TIEMQKKLNTIQMLLENITFSVQKPITTSNLT----KIRS----ILSYLGITSETAYTDIL 171
      +++ L IQ L N P ++ I+S +LS LGI E+ D++
Sbjct: 120  HTRLEQSLYDIQKSLSNVLQSGIPTQVNDQVHDDSIKSYGQYLLSELGIAGESGSKULM 179

Query: 172  NICHLLLKQELMF-----AQDFQKELSIDE-----HQKKIILQRIRRAVKK 213
      NI L E + A D ++L+ ++ + K QR+RRAV +
Sbjct: 180  NILMFLYTYEKEYSFEKGFALKDIFQLASEKLGDAADERDVRVRVKAQQRVRAVYQ 239

Query: 214  AMINMAHYIDDFENSLTLQYANALFGPQNTHNEAQLIQGK--SMYGGKISLGHFDEL 270
      ++ +AA L + DF N +YA+ F F ++ ++ + S +I++K F L
Sbjct: 240  SLHWASLGLIDFNSNPKPEYASHFFDFSVVRSSMTLKHNETSSSYTSARINVKFKTAL 299

Query: 271  ILQSK 275
      ++K
Sbjct: 300  YTEAK 304

```

30 There is homology to SEQ ID 460.

A related GBS gene <SEQ ID 8723> and protein <SEQ ID 8724> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1 Crend: 8
MoG: Discrim Score: -7.05
GvH: Signal Score (-7.5): -6.58
      Possible site: 61
>>> Seems to have no N-terminal signal sequence
ALOM program count: 1 value: -5.52 threshold: 0.0
      INTEGRAL Likelihood = -5.52 Transmembrane 49 - 65 ( 46 - 66)
      PERIPHERAL Likelihood = 7.37 155
      modified ALOM score: 1.60

*** Reasoning Step: 3

----- Final Results -----
      bacterial membrane --- Certainty=0.3208 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

50 The protein has homology with the following sequences in the databases:

```

ORF00604 (307 - 1125 of 1431)
EGAD|137180|146289 (3 - 304 of 310) hypothetical protein [Bacillus cereus]
GP|1769946|emb|CRA67094.1|X98455 orf1 [Bacillus cereus]
%Match = 12.7
%identity = 34.1 %similarity = 53.0
Matches = 95 Mismatches = 123 Conservative Sub.s = 53

168      198      228      258      288      318      348      375
*C*W*YLSRNRIRPRAYGNRAISRNNCLSSAKWNNIYVIP*KSI*VRR*YVKFYIIDDPTITMILQDIEE-DFN
      :||:|||| :| |||:|:
MFYYIVDDEVFRSMLSQIIEEDGLG
      10      20

```


-1175-

```

405      435      465      495      525      555      585      615
NTVVRNVNYSKAYNELLDADVIVLIDLLNPILDGVTLVQKIYKQSDLPKFMISQVKINDLRQAYKAGIEFFINKPI
:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
5 EVIGSESDGAFVEASQLNKKYVDILFIDLLMPMRDGIETVRIH-ASSFPGKIIMISQVSKQLIGRAYTLGVEYITIKPL
      40      50      60      70      80      90      100

645      675      705      753      771      801      831
NILEVSVVVRKVTDTIEQMOKLNTIONLLENTPSYQKP---ITTSNLTKI---RSLISYLGITSETAYTDLINICELL
|   |   |   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
10 NKIEVSVVVRKVIERRIRLSIYDIQKSLNNVPQWKGPKMRSETVOERKKISDSGRFLAELGAGENSDDLMSLEYL
      120      130      140      150      160      170      180

861      894      924      954      984      1014
LKQRLNFAQDFQKELSID-----RHQKIKILQRIKRAVKKAMINMAHLYIDDFENELTLQYANAL
|   |   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
15 YGGE-KAQTFEPGPPALKDIFHQITLKKLGRISADADIEKKKASBQVRERAIYQSLNHLASLGLTDFSNKPFESYAKKF
      200      210      220      230      240      250      260

1071      1095      1125      1155      1185      1215      1245
FGPGNIHNE-AQLIQGSKSYGGKISL--KHFFDELILQSKTF*DLFKHGLIYNNHFKTFLFIDNLQTFCLPQGVCFPF*F
|   |   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
20 FDTFVVRKNTMTKDGVAISGHIRINTKKFIVLYFEAKRLMEIE
      280      290      300      310

```

SEQ ID 8724 (GBS356) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 3; MW 34kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 8; MW 59kDa).

GBS356-GST was purified as shown in Figure 216, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1051

A DNA sequence (GBSx1123) was identified in *S.agalactiae* <SEQ ID 3243> which encodes the amino acid sequence <SEQ ID 3244>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have a cleavable N-term signal seq.
35 INTEGRAL Likelihood = -6.48 Transmembrane 149 - 165 ( 147 - 172)
INTEGRAL Likelihood = -5.20 Transmembrane 37 - 53 ( 29 - 55)
INTEGRAL Likelihood = -2.50 Transmembrane 126 - 142 ( 126 - 142)
INTEGRAL Likelihood = -2.13 Transmembrane 62 - 78 ( 60 - 78)
40 INTEGRAL Likelihood = -0.64 Transmembrane 314 - 330 ( 314 - 330)
INTEGRAL Likelihood = -0.11 Transmembrane 89 - 105 ( 89 - 105)

----- Final Results -----
bacterial membrane --- Certainty=0.3590 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
45 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB06435 GB:AP001516 two-component sensor histidine kinase
[Bacillus halodurans]
50 Identites = 118/427 (27%), Positives = 199/427 (45%), Gaps = 25/427 (5%)

Query: 10 LERRQRIIISAIAIA-LAAQINISILADGFMITLSLFLPVLVFNDDINPILCLGITF 68
L + II+S + A +A +IN + + F +L I +FL F + I+
Sbjct: 7 LSKDYMILGSKLAFAPAGKINFPTVNETRVSGLGPPIFFLFLPLNTAIVGVFTAI 66

55 Query: 69 ASPIFRGIIISLAGEARHQIEFVLATMAYFCYIGTFTYTYMRSYRNKGITFFFSIII 128
A +FR + + + + E FT Y + F R + F II
Sbjct: 67 AVVVFRVFLDTLHADFPYWDSPETIHYPTFFFTYSLLSLAKVQRPHQPLIIFLGII 126

```

-1176-

Query: 129 CDYFANLVEISFLIKNFYITITIPA-TLPAIALIRAFISCAVAYTYSYLLQKD---D 184
 + A+ R E+ ++ + + + ++F I L+ S V +S + L + + +
 5 Sbjet: 127 IEILADTAE--FIAQYFAPGVWVTKDSIFQILLIAPSHSFIVLGVPSMMKLKYEGRSRELE 184

Query: 185 HERRYYYPWMSAVKSVVYPMQKNIIRIKINIMKNAYLLDKELSKY---HLPKRYQLS- 240
 +R + + + S + E ++K + E+I + L +B+ H+ + HL
 Sbjet: 185 IRKRNEHMLLIISNLYESVHLKCTIQNSEDITSKVFGLYREMKRLQSEHMQVNFHEK 244

10 Query: 241 ----LIDSRDVHEVQKDYQNIILKGLTYPVSNIESTMALKDI PQIVLSYTRS---IIQF 292
 L+IS +VHE+KQD Q I GL S NES + +I QI+ R+ Q
 Sbjet: 245 ISKRLLSISGEVHEIKKURKIFAGLSKLIS--NBSYVDYITIGQIKINIVKTEKYAQL 302

15 Query: 293 RHQDIILLENKONLIISNYLLTITISNIVNAVEAIDKQKQKTSVHTESIEDSPKIE 352
 ++I + + + Y L+II+N+V NAVEAID KG +++ + L ++
 Sbjet: 303 LGKEIDPHYISQSEHPYHYTHLSITNNLVANAVEAIDG--KGLMTIRVKALQTVTEFR 360

20 Query: 353 ISDNFGQIPDRMKHMHIFKPGFSTKFDANGDIYRIGLSHVRLIMEQYQOQTITVCPNQ-P 411
 I D+GQIPDK + +IFKPGF++KPD G GIGL++V M ++ GT+ Q
 Sbjet: 361 IEDDQFGIPDKRALIFKPGFSTKFDHTGKPSGIGLTVVHD-MVDLQSTVVYERQCG 419

Query: 412 NGTFVTL 418
 G+ FT+
 Sbjet: 420 KGSVPTI 426

25

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1052

30 A DNA sequence (GBSx1124) was identified in *S. agalactiae* <SEQ ID 3245> which encodes the amino acid sequence <SEQ ID 3246>. This protein is predicted to be ornithine carbamoyltransferase Otc6850 (argF). Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 35 INTEGRAL Likelihood = -0.64 Transmembrane 171 - 187 (171 - 187)

----- Final Results -----
 bacterial membrane --- Certainty=0.1255 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 40 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB75986 GB:A2272085 ornithine carbamoyltransferase
 [Staphylococcus aureus]
 Identities = 264/332 (79%), Positives = 292/332 (87%)
 45 Query: 1 MKNLNNRSPFLTLDPSFAVEFLLKSLSDLKRAKYAGIEQOKLVGNKIALIFEKDSITR 60
 MKNLNNRSPFLTLDPS EVRFLL LGRDLKRAKY G E+ L KNIAL+FEKDSITR
 Sbjet: 1 MKNLNNRSPFLTLDPSRQVEFLLTSEDLKRAKYIGTEKMLKNKIALIFEKDSITR 60

50 Query: 61 CAFEVAARDQGANVYTLGPTGSGMGKKRISKITARVLGGMYDGIERYGFSQRTVETLAEF 120
 CAFEVAARDQGA+VYTLGPTGSGMGKKR+KIDARVLGGMYDGIERYGFSQ TVEITLAE+
 Sbjet: 61 CAFEVAARDQGANVYTLGPTGSGMGKKRITKITARVLGGMYDGIERYGFSQRTVETLAEY 120

55 Query: 121 SGVPVWNGLTDHPTQVLADFLTAKGLKPKYDIFFTYVSGDGRNNVANALMGASIVG 180
 SGVPVWNGLTD DHPTQVLADFLTAK L K Y DI FTYVSGDGRNNVANALM GA+I+G
 Sbjet: 121 SGVPVWNGLTDHPTQVLADFLTAKSVLKQYADINFTYVSGDGRNNVANALMGGAING 180

Query: 181 MTYHLVCPKELEPDPELLSKQBIKATKTGASIEITADIAEGVRDSVLYTVWVSMGEPD 240
 M +HLVCPKEL P ELL++C+ IA G +I IT DI +GV+ SDV+YTDVWVSMGEPD
 60 Sbjet: 181 MNFHLVCPKEIMPTDRLNRCNKNTAENNGNILLITODIDQGVKSGSVIYTVWVSMGEPD 240

-1177-

Query: 241 EVWKERIALLEPYRITQEMLANWENPVI FEHCLPSFHNDTKVGYDIYKYLKMEVS 300
 EVWKER+ LL+PY++ +EM++ T NPNVIFEHCLPSFHN DTK+G I+EKYG++EMEV+
 Sbjct: 241 EVWKERLELLKPYQVHKSMMDRTGNPNVIFEHCLPSFHNADTKIGQQIFEKYGIREMEVT 300

5 Query: 301 DEVFEGHMSVVPQEAENRMTITKAVNWATLG 332
 DEVFE SUVFQEAENRMTITKAVNWATLG+
 Sbjct: 301 DEVFESKASVVPQEAENRMTITKAVNWATLGE 332

There is also homology to SEQ ID 3118.

- 10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1053

A DNA sequence (GBSx1126) was identified in *S. galactiae* <SEQ ID 3247> which encodes the amino acid sequence <SEQ ID 3248>. This protein is predicted to be carbamate kinase (b2874). Analysis of this protein sequence reveals the following:

Possible site: 53
 >> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.48 Transmembrane 214 - 230 (214 - 230)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.1192 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 25 The protein has homology with the following sequences in the GENPEPT database.

>CP:CAA66367 GB:X97768 carbamate kinase [Clostridium perfringens]
 Identities = 162/313 (51%), Positives = 207/313 (65%), Gaps = 7/313 (2%)

Query: 3 KIVVALQGNAL-----QNSPEQLRLVKHTAKSLVALIKKGHEIVSHNGPQVGAINLG 57
 KIV+ALG NAL S E QL + TA S+ LI+ GHE+ + HGNGPQVG I
 Sbjct: 2 KIVLALGENALQKDSKDKSAGGQLETCTQTALISVALIEDGHEVSIVHGKGPQVQILAS 61

Query: 58 MNFAAESGGQTN-PPFPGCGWSSQYIGYHLQSSLINELRQBGINKEVATITITQEVDES 116
 + A + G PFP GA S+GYIGYHLQ ++ EL + GI K V TI TQ+ VD++
 Sbjct: 62 IELAHQVNGNPLPPFDVVGAPSEGYIGYHLQNTIREELLKRGTEKSVDTITTTQVIVDKN 121

Query: 117 DQAFSAPTKPIGTFFYDKETSEKIAIEKGYTFVEDAGRGYRRVVASPEPKKIEINSIKTL 176
 D P+ PTKPIG+PY KE +EK+ +KGVT EDAGRGYRRVVASP+ I+E +IKT+
 Sbjct: 122 DPGFTNPTKPTGSPYTKKEEARKLEKDKGYTMKEDAGRGYRRVVASPKPVDIVEKEAIKTM 181

Query: 177 IENDLWLAGGGGGIPVINKGS-YEGIAAVIDKDKSALLAGELADOLILITAVDQVYTT 235
 +++ +VIA GGGGIPV+ G BG+ AVIDKD ++ LA L AD L+ILTAVO V
 Sbjct: 182 VDSGPVIVACGGGIPVVEDGRLGVPVAVIDKDPAAKILABELADALLILITAVDRVCV 241

Query: 236 QPGKENQKALTEVENRQMDIVYNGQFPAKGSMLPKVIAWMSFLDENKGTALITSLNGL 295
 F K +QKAL E+N ++ Y+ +G+PA GSMLPKV AC F+ K A+I SL +
 Sbjct: 242 NFNKPDQKALKEINLEDEKVIIEGQFAPGSMLPKVIRACKKFLVSGDKKVALIASLTNAK 301

Query: 296 DALDGKLGTRITK 308
 AL G+ GT+T K
 Sbjct: 302 AALRGESGTYKIVK 314

There is also homology to SEQ ID 3110.

- 55 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1054

A DNA sequence (GBSx1127) was identified in *S.agalactiae* <SEQ ID 3249> which encodes the amino acid sequence <SEQ ID 3250>. Analysis of this protein sequence reveals the following:

```

Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3558 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1055

A DNA sequence (GBSx1128) was identified in *S.agalactiae* <SEQ ID 3251> which encodes the amino acid sequence <SEQ ID 3252>. This protein is predicted to be a transmembrane protein (b2298). Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -13.11 Transmembrane 413 - 429 ( 405 - 440)
INTEGRAL Likelihood = -9.61 Transmembrane 498 - 514 ( 489 - 516)
INTEGRAL Likelihood = -9.45 Transmembrane 165 - 181 ( 161 - 185)
INTEGRAL Likelihood = -8.07 Transmembrane 127 - 143 ( 122 - 146)
INTEGRAL Likelihood = -7.22 Transmembrane 308 - 324 ( 306 - 326)
INTEGRAL Likelihood = -5.57 Transmembrane 334 - 350 ( 330 - 357)
INTEGRAL Likelihood = -4.51 Transmembrane 194 - 210 ( 193 - 217)
INTEGRAL Likelihood = -3.82 Transmembrane 372 - 388 ( 371 - 390)
INTEGRAL Likelihood = -1.22 Transmembrane 250 - 266 ( 250 - 268)
INTEGRAL Likelihood = -0.80 Transmembrane 468 - 484 ( 468 - 484)
INTEGRAL Likelihood = -0.32 Transmembrane 436 - 452 ( 436 - 452)

----- Final Results -----
bacterial membrane --- Certainty=0.6243 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC22251 GB:U32741 conserved hypothetical transmembrane protein
[Haemophilus influenzae Rd]
Identities = 303/506 (59%), Positives = 389/506 (77%), Gaps = 6/506 (1%)

Query: 10 NKRSGRFRMPGAPTILFILITIPSVLITWMLPAGSYSKLQFDTASSKLVTVDNPKTVEVP 69
      +R+ K F P AFTILE + I + V TW IF+GSYSKL +++ + VV P
Sbjct: 4 SKKKKTTFNPSAFTILFALILAVGLTAVIPSGSYSKLTYNSTINVPVVKAYGVDDKTYP 63

Query: 70 ATQTQLDRKNNVKIKIKEFTSGAISKPSVNTYKRLQKNPAGISGVITSMVNGTIEAVDI 129
      AT LD +N+KIK+ FT G I KP+++P TY+R++Q+ GI +T SMV GTIEAVD+
Sbjct: 64 ATTDLDLNLNLIKILSNFTBQVKKPPIALPGTYQRVQBQKHGIEDTKSMVGTIEAVDV 123

Query: 130 MYPFIVLQGMIGVVRKSGAFESGLIALTKTKGRFLLIPLVSLIMVIGTLGJIEEAV 189
      MVFI VLQGMIGV+ ++G+F +GL+AL KTKTG EF ++F VS+LMVIGQT CSJIEEAV
Sbjct: 124 MYPFIVLQGMIGVINRTGSPNAGLMALVKTKGRNEFFIVPCVSLIMVIGTTCJIEEAV 183

Query: 190 AFYPILVPIFIAMGYDSICVGAIFLASSVGTFSFTINPFSVIASNAAGISFTBGLSWR 249

```

-1179-

AFYPLVP+FLA+GYD+I+CVCAIFLA+S+GT+ESTINPFS VIASNAAGI FTEG+ +R
 Sbjct: 184 APTPLVPVFLALGYDAIVCVCAIFLAASMTAFSTINPFSVIASNAAGIQFTEIGR 243

5 Query: 250 TACCAIGALFVVVLLHWYAKKIKANPEFSYSYEDRVEFMAKNGMTIN-HPSSLFTTRKQI 308
 G + GA V+ YL+WY KKIK+A+ FSY+Y+DR EF ++ + +T F+ R+K+
 Sbjct: 244 ALGLVLGATCVIATLYWYCKIKADPFSSTYTDREERFQRKMKNFDTNTIPPSARRKL 303

10 Query: 309 ILSLFLVISEFLMVGVMVSGQWFFPMASFLAIIIMFLPATGANGIGERDVDFEVNG 368
 IL+LW ISF+M+MGVM GMMF MA+SFLAIIIMF+ +G+ E+D++ F G
 Sbjct: 304 ILTLPCISFPDIMGVMVQGMFPQMAASFLAIIIMFI----SGLSKDIMEFTEG 358

15 Query: 369 ASSLVGVSLIIGLARG+NIILSQYISDTMLYTSKASHVSGSVIIVMMETIYFVLGV 428
 AS LVGVSLIIGLARG+NI+L Q3 ISDT+L 8 + 8 + GSVFI+ ++ LG +
 Sbjct: 359 ASLVGVSLIIGLARGVNLVLEQGMISDTILDYMGVNVSGMPGVSFILGQLVVFIFLGLI 418

20 Query: 429 VPSSGLAVLSMPILAPLADTVGIPRVSVMAYQFGQYAMLFAPLGLVATLQMLDMKY 488
 VPSSGLAVLSMPI+APLAD+VGIPR +VY AV +QYAMLFAPLGLV+ TLQML ++
 Sbjct: 419 VPSSGLAVLSMPINAPLADSVGIPRDIIVSVAYNWQYAMLFAPLGLVIVLQMLQIFP 478

25 Query: 489 SHNLKFWFVVLFLILFGGLILVQV 514
 W+KFW P++ LL+ G LV+QV
 Sbjct: 479 DRNVKFWFMIGCLILGSLILVQV 504

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3253> which encodes the amino acid sequence <SEQ ID 3254>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have a cleavable N-term signal seq.

30 INTEGRAL Likelihood = -13.21 Transmembrane 479 - 495 (472 - 496)
 INTEGRAL Likelihood = -10.24 Transmembrane 261 - 277 (258 - 280)
 INTEGRAL Likelihood = -9.24 Transmembrane 153 - 169 (142 - 180)
 INTEGRAL Likelihood = -7.17 Transmembrane 393 - 409 (391 - 411)
 INTEGRAL Likelihood = -6.00 Transmembrane 81 - 97 (78 - 99)
 INTEGRAL Likelihood = -5.95 Transmembrane 318 - 334 (314 - 338)
 INTEGRAL Likelihood = -3.77 Transmembrane 352 - 368 (352 - 369)
 35 INTEGRAL Likelihood = -2.56 Transmembrane 120 - 136 (119 - 138)
 INTEGRAL Likelihood = -0.32 Transmembrane 204 - 220 (204 - 220)

----- Final Results -----
 40 bacterial membrane --- Certainty=0.6286(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:ABE94000 GB:AP008219 unknown [Borrelia afzelii]
 Identities = 174/496 (35%), Positives = 306/496 (61%), Gaps = 37/496 (7%)

45 Query: 10 RIPSSTYVLPIIILAMVLTMPIPAGAYETAK ---GGG-----VISCTYKTVAGNQGF 61
 ++PS+T++F +I + +L+ IPAG ++ G G ++GTY+T+ P+GF
 Sbjct: 3 KMPSSFTIIPSLIVFVTLITYIPACKFDKSPHQIGDGPKRRIIVAGIYQITDRGFRGL 62

50 Query: 62 DILMAPVGMGLVGEVDGAIQVSEFPIIAGVGGLGVNKGALDGTIASVVRHKKRKMFL 121
 +M + M +G + A +V F+L+VGG G++ KTG+D GI S++K ++K+L
 Sbjct: 63 HPIIMTILAMS -KOMEHAAEVLIIPVLVIGGAYGIDMKTGADAGIYSLIKLGHKRL 120

55 Query: 122 IAILPLPALGCTTYCAGETMAFYPLIPVIMVIAVGFSDIVAVAILIGSQIGCLASTIN 181
 I +L+ +P+GCT GM ERT+ FY ++IP+++A+G+D+V VAIL +G+ +G +ST+N
 Sbjct: 121 IELLMFIPSIGCTYVCMSEBETLPFYVMIPLIVAGYDNNVGVAILAGGVGTINASTV 180

60 Query: 182 PFATGVAADAGVSAIDKIMKVIQWVILVGMISIMFYVNYASKIREDPSKSLVADKEKH 241
 PFATG+A+ A +S+ DG +R++ + I + ++I +V YAS+I+D+DSKSLV K+ EH
 Sbjct: 181 PFATGIASIASISLQDGSFRIVLYPISILVAILTYCVYASIKKDKPSKSLVSKQEH 240

65 Query: 242 KFLP-QLNCGSDILNKRGRNVITPILTFVIMILSLPWEDGKIFPFINIVMLTMMPL 300
 + F + + S ED NV TF ++ L+ PG I + + + L
 Sbjct: 241 YQFVKNISKED-----NVQNTLEPTFARKLVILL-----FGFW---ILFLPSPVIVQL 286

-1180-

Query: 301 GGVIGKTMGAGFTWYFPEITMLFIMMGVLVAIVYRMSEDFPSSFLTGTAGEFLGVAMICA 360
 G W+ E+TML++ ++ A + R+ E + + +F+ G+ + A+I
 5 Sbjct: 287 G-----WMMQEMIMLYLGVAISAFICRLGESEMDAFVKGSSSLITPAALLIG 334

Query: 361 IARGIQVIMNGGMITATILHLGFTSLSGLSQVFLIAYIFVLMSSFLIPSTSLAGATM 420
 +ARG+ ++ + G+ITAT+L+ L L F+IL I + + F++PS+G A TM
 Sbjct: 335 IARGVMVIMCDDGLITATMLNADTNFLYNLRPFPFIILNEIIQIFIGFIVPSSSGHASLTM 394

10 Query: 421 GIMAPLQCFSNVPAHLVITAFQASGILNMISPTSIAVMGALALGRVDLTWNKPIGKFI 480
 IMAPL F ++ V+ A Q++SG++N+I+PIS ++M L + ++ GTW+KF+
 Sbjct: 395 PIMAPLADFLSIRSSSVIAMQTSGLINLITPTSGVMVAVLGISKLSYGTWKFVPLEF 454

15 Query: 481 VMVMLSVLLLVVATF 496
 ++ +S+L+++ +
 Sbjct: 455 IIEFFISILLVIANVY 470

An alignment of the GAS and GBS proteins is shown below.

Identities = 158/542 (29%), Positives = 274/542 (50%), Gaps = 92/542 (16%)

20 Query: 11 KRSKGFRMPGAPFTILFILTIPSLATWNIIPAGSYKLGQFDTASGKLIVTPDGKTVHVA 70
 ++ +GPR+P ++T+LFI+ + TW+IPAG+Y +TA
 Sbjct: 4 EKKRGFRIPSSYTVLFIILAMAVLWTFIPAGY-----ETARG-----42

25 Query: 71 TQTQLDKMNVKIKIKFTSGAISKPVSVPNTYKRLKQNPAGISGVTTSMVG-----TI 124
 G IS TYK+ NP G + + V G T
 Sbjct: 43 -----GGVIS-----GTYKTVASNPQGFIDMAPVRMGLGVETD 78

30 Query: 125 EAVDIMVFIMVLGEMIGVVRKSGAFESGLLALTKTKTGREFLLIFLVNLSLMLVGLTLOGI 184
 A+ + FI+++GG+GVV K+GA ++G+ ++ +K KGRE +LI ++ L LGST G+
 Sbjct: 79 GAIQVSFFILMVGGFLGVNKGALDTCIASVVRNKGREKMLIALILPLFALGFTTYGN 138

35 Query: 185 EEEAVAFYPLVPLFPLMGVDSICVGAIFLASSVGTSTFINPPSSVIANAGISPT 244
 EE +AFVP+L+P+ +A+G+DSI+ V I + S +G STINFP++ +A++AAG+S +
 Sbjct: 139 GEETMAFYPLLIPLVMAVGFDSIVAVAILLIGSQIGCLASTINPFTGVAADAAGVSIA 198

40 Query: 245 GLSWRTAGCIAGAIPTVVVYLHWYAKKIKANPEFSYSYEDRVFNKMGMITNHTPSLPTI 304
 G+ VR + + +++ Ya KI+ +P S D+ E + + N L
 Sbjct: 199 GMIWRVQWVLVGMISINFYNYASKIEEDPSKSL-VADKBEHKGSLFOLQNSGEEL-NK 256

45 Query: 305 RQKILLSLFLVIFPLMV-----GVMG-----GVMF 331
 RQ+ +L+P ++F +M+ W GV+ + W+P
 Sbjct: 257 RQENVLTIFLTFVIMLSLIPWDFGKIFFTININTWLTMPILGLGVIGKTMGAGFTWYF 316

50 Query: 332 PTHASSFLAITIIMFLATGANGIGERDVVDFVNGASSIVGSLVLIIGLARGINIILSQ 391
 P + F+ + +++ + + E D F+ GA +GV++I +ARGI +I++
 Sbjct: 317 PEITMLFIMMGVLVAIVYR-----MSSEDFSSFLTGTAGEFLGVAMI CAIARGIQVING 371

Query: 392 GYISDTMLNTASKLASHVSGSVFIIVMMFIYFVLGVFPSSSGSLAVLMPILAPLADTVG 451
 G I+ T+L+ S +S VF+I+ Y + F++PS+GLA +M I+APL
 Sbjct: 372 GMITATILHLGFTSLSGLSQVFLIAYIFVLMSSFLIPSTSLAGATMIGMAPLQPSN 431

55 Query: 452 IPRSVVMAYQFQYAMLEFLAPT-GLVMATLQMLDMKYSHWLKFMVFPVFLIFLIPGGLLV 512
 +P +V+ A+Q + ++PT +VM L + + W KP+ ++ +++ LLV+
 Sbjct: 432 VPAHLVITAFQASGILNMISPTSIAVMGALALGRVDLTWNKPIGKFIIVMMLSVLLLVY 493

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1056

60 A DNA sequence (GBSx1129) was identified in *S. agalactiae* <SEQ ID 3255> which encodes the amino acid sequence <SEQ ID 3256>. Analysis of this protein sequence reveals the following:

Possible site: 46

-1181-

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood =-10.83 Transmembrane 25 - 41 (18 - 47)
 INTEGRAL Likelihood =-10.46 Transmembrane 153 - 169 (148 - 176)

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the GENPEPT database.
 >GP:CAB13183 GB:Z99110 similar to two-component sensor histidine
 Kinase [YkoG] [Bacillus subtilis]
 Identities = 119/446 (26%), Positives = 212/446 (46%), Gaps = 18/446 (4%)

15 Query: 17 TQITLWYSSFFILVIGVLGVSFFISKSIABNKSKNLEAKAVQMSQMLAKGHRYEAFED 76
 T+I L+ S + IL+I V + I S +K L + +++AL
 Sbjct: 5 TKIHLVTSISLLILLLVHTVAVLIPSSALTSKDAVLADETNDINAEALRAAEETGVALQ 64

20 Query: 77 GIFYSVYDQNGKV-IYSGFPGFKRDLQHGHKKLKLFSMEN-----RTFCYVDI 127
 + + NG V + +G K + LS S E + F +
 Sbjct: 65 DMLQAYLPANGMVVYVNGQKAVMTITTKSKAYKDFPLSFIRSGETADVVRKPDGKLFAEAAV 124

25 Query: 128 PISGKNQNLRAIRTVDRDLKQITELLFSLGIVLFLMLIITVG---GYLIKRTFRPIQ 183
 P+ + + + + V+RL+ E LF L I+L + + G L+ +R PI+
 Sbjct: 125 FVINTDQGVVSLGVLRLENT-EESLFLKILIIAASAAVCIASFAGSLLARRIINPIR 183

30 Query: 184 EITETAQFITQNYEDYTKRIITKNENELTELAAVINTMLASIESSFVREKQFNNOVSHEL 243
 + T + I + + + + + + + EL + + N M + + + + + QF D SHEL
 Sbjct: 164 RLMTMKDIQRKEFKTISLGQSNDELYQMLTFNEMAMMKHEHYDQQQVQDASEL 243

35 Query: 244 RTPVTILSSSEYKGYAENLSA-KESFEVIHQSLGSKKLVLQELLETKAENPLSIQL 302
 +TP+T+I S S K + E +ES E IH + + + MKL CLL L K+ L + L
 Sbjct: 244 RTPVTIIESYSLMKRWGAKKFEVLEESIEATHSBAVMKKLITVQLLALAKSHQGLEVDL 303

40 Query: 303 EPLNFSIMMKQLVSDSSRLDNTPIHLDSQIEDDLWIIQQITLKLRLPDNLFSNAIKFTN 362
 + + + I + V + + + I L+ + + L + + +K+L L NAIK+ +
 Sbjct: 304 KTI DL- IKAARAVIQTLQSVYQRDILLET- KESLLVKADEEIKQLLTILLDIAIKYSE 361

45 Query: 363 NHISILRSDNOIVFSIKINGLGISVDDQSKIWRFTYVDSARTKDSQSGIGLGLSLVK 422
 I +S + + S+D G+GI + ++ RFT+ D AR + + G GLGLS+ K
 Sbjct: 362 KPIEMSAQTRNGRPFLSVRDEGIGIPEHIIPLHFERFYRDEARNKKT-GGTGLGLSLAK 420

Query: 423 QIATIHRAKIWDVSKPDGGSQPTTF 448
 QIA H + + V SKP G+ T+ F
 Sbjct: 421 QIADHIGILSVKSKFGQGTAVTMQF 446

There is also homology to SEQ ID 1178.

SEQ ID 3256 (GBS77) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 21 (lane 2; MW 78.5kDa) and in Figure 28 (lane 2; MW 78.5kDa).

50 GBS77-GST was purified as shown in Figure 195, lane 4.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1057

A DNA sequence (GBSx1130) was identified in *S. agalactiae* <SEQ ID 3257> which encodes the amino acid sequence <SEQ ID 3258>. This protein is predicted to be CopR protein (trcA). Analysis of this protein sequence reveals the following:

Possible site: 33

-1182-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3963 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CA07978 GB:AJ278983 CopR protein [Ralstonia metallidurans]
 Identities = 102/221 (46%), Positives = 145/221 (65%)

Query: 1 MKILVVEDEFDLNRISIVKLLKKQHYSDASNGERAIQFVSVAKYDVIIILVMMPKMDGF 60
 MK+LVVEDE + + L + + VD +NG + F YD+IILDMV+P +DG+
 15 Sbjct: 1 MKILVVEDEVKTEGYLRQGLTEAGFVVDLVANGLDQGHFAVNETYDLIILVMLPDVDGN 60

Query: 61 TFLKLLRNKGSQVSIIMLTARDAVEDRIAGLDPGADDYLVKPFEGEIMARIMARIMLRAN 120
 L +R G+ V +L LTARD+V DR+ GL+ GADYLVKPPF F EL+AR+R +LRK
 15 Sbjct: 61 HILHAIIRASGNVFPVLPILTARDGVADRVRGLELGADYLVKPPFAPSEILARVITLRRGA 120

Query: 121 RQVSSDDIQIDITINLSTKQVWRNDNLIDLITAKRYEVLKYLAHRHJQVLSHQIRESHVN 180
 Q++ D IQ+ D+ ++LS ++ R I LT+KE+ +LE AR R +VL R I VM
 20 Sbjct: 121 VQLAMDRICQVADLILDLISRRRRASRGRRITLTSKEFALLELPARRRGVEVLPESLTSQVM 180

Query: 181 DYDYGYGSNIIDVLIHNLRRKLDNNRDSGLIKTKRGLGYVI 221
 D ++ +SN+IDV I+ LR K+D+ + LI+T RQ+GYV+
 25 Sbjct: 181 DMNFDSDSNVIDVAIRLRARAKIDGFEVKLIQTVRNGYVL 221

There is also homology to SEQ ID 3260.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 30 vaccines or diagnostics.

Example 1058

A DNA sequence (GBSx1131) was identified in *S.galactiae* <SEQ ID 3261> which encodes the amino acid sequence <SEQ ID 3262>. Analysis of this protein sequence reveals the following:

35 Possible site: 40
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.45 Transmembrane 18 - 34 (16 - 36)

----- Final Results -----

40 bacterial membrane --- Certainty=0.2381 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10281> which encodes amino acid sequence <SEQ ID 10282> was also identified.

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3262 (GBS78) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 4; MW 23.8kDa).

The GBS78-GST fusion product was purified (Figure 194, lane 4) and used to immunise mice. The
 50 resulting antiserum was used for FACS (Figure 317), which confirmed that the protein is immunoreactive on GBS bacteria.

-1183-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1059

A DNA sequence (GBSx1132) was identified in *S.agalactiae* <SEQ ID 3263> which encodes the amino acid sequence <SEQ ID 3264>. Analysis of this protein sequence reveals the following:

```

Possible site: 36
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -11.04    Transmembrane    15 - 31 ( 6 - 35)
    INTEGRAL    Likelihood = -1.28    Transmembrane    51 - 67 ( 51 - 67)

----- Final Results -----
    bacterial membrane --- Certainty=0.5416 (Affirmative) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3264 (GBS79) was expressed in *E.coli* as a GST-fusion product. GBS79d was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 154 (lane 17 & 18; MW 51kDa), in Figure 155 (lane 17; MW 51kDa) and in Figure 187 (lane 13; MW 51kDa). GBS79d was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 2-4; MW 26kDa) and in Figure 183 (lane 5; MW 26kDa). Purified GBS79d-GST is shown in Figure 243, lane 2.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1060

A DNA sequence (GBSx1133) was identified in *S.agalactiae* <SEQ ID 3265> which encodes the amino acid sequence <SEQ ID 3266>. Analysis of this protein sequence reveals the following:

```

Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
    bacterial cytoplasm --- Certainty=0.5326 (Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10279> which encodes amino acid sequence <SEQ ID 10280> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:ANG20974 GB:AE005164 Vng6345c [Halobacterium sp. NRC-1]
Identities = 97/358 (27%), Positives = 163/358 (45%), Gaps = 20/358 (5%)

Query: 35 DPQIKITIRANIAIGTYEGFLESINIPMLISPLLSQEAVALSKLBGTHATLKDLLNYS 94
          D + A +G G + P +L + LL +EA+ S++EG L + R
Sbjct: 70 DDDFYETLADATFNLGKSGVSLDLFPFVLVTSLLRKEAMESARTIGADVDVYDALYSLS 129

Query: 95 AGNKVDIRDELHEIIT-----NYRKALFYALENISTINNIDSKPLSLNRITKEHKIL 148
          D RDE E + R+ L Y I+ +D+ G L+ ++ ++H+ L

```

-1184-

Sbjct: 130 T-RTFDEGRDEPSETTAAAEFKDTREVLANYETAVKEGIDALDA-GEELNVELLADLHETFL 187
 Query: 149 LINV---RGSSKNPGNFKRSQNYIGSVSSISYTPVPAKTEPYMSNLECYIHVD-LDLIL 204
 L V R + G++K + NY+G + P + M L Y L
 Sbjct: 188 LTGVFDERVDITVIGDYKINPHYLGD-----FLPAPGAVEDLMDGLPTTYRTGSSYHPL 242
 Query: 205 VQSALIHQFEMIHPPEDGNGRIGRLILFLFYQRLLSVPTVMSYPERDRSLYISHL 264
 V A+ H QFE IHP+ DGNR+GRLLI L LY +LL P Y+S Y R+++ Y+ +
 Sbjct: 243 VDIALFTYQFETTHPYGDGNGRLRLITLQLYDADLLERPNLYSEYLRNKKTTVERM 302
 Query: 265 ENISKGNWKNWDFEYLYLGVLASBSTKKAQDILSLYNIMKEQVTPKINSVSGIQLDLP 324
 + W+ W +++EG+ A ES +++ L + + K + + Q
 Sbjct: 303 EGVRFHGEWAMLSFFPIGSIARQAHSVERTRALADLRREYHEYGGKATTN--QLAVT 360
 Query: 325 IFSAPIFKARQVSEHLKLSKRTTYTLANKLIDBGL-STDNAQRNRTTYQFOLLISIV 381
 +F P ++ V I + T +H+L+EG L RN+ Y ++ I++
 Sbjct: 361 LFEQFYITSTKVQRLFDIEQSTASPAINELVNBSGILKEVPRHGRNKEYRAREIFEILE 418

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1061

A DNA sequence (GBSx1134) was identified in *S.agalactiae* <SEQ ID 3267> which encodes the amino acid sequence <SEQ ID 3268>. Analysis of this protein sequence reveals the following:

- 25 Possible site: 47
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.4370 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 RGD motif : 46-48

- 35 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3268 (GBS299) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 2; MW 62.2kDa) and in Figure 60 (lane 4; MW 62.2kDa).

GBS299-GST was purified as shown in Figure 207 (lane 4) and Figure 225 (lanes 2-3).

- 40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1062

A DNA sequence (GBSx1135) was identified in *S.agalactiae* <SEQ ID 3269> which encodes the amino acid sequence <SEQ ID 3270>. Analysis of this protein sequence reveals the following:

- 45 Possible site: 37
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.4176 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1063

A DNA sequence (GBSx1136) was identified in *S.agalactiae* <SEQ ID 3271> which encodes the amino acid sequence <SEQ ID 3272>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1789 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1064

A DNA sequence (GBSx1137) was identified in *S.agalactiae* <SEQ ID 3273> which encodes the amino acid sequence <SEQ ID 3274>. Analysis of this protein sequence reveals the following:

```

Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3748 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1065

A DNA sequence (GBSx1138) was identified in *S.agalactiae* <SEQ ID 3275> which encodes the amino acid sequence <SEQ ID 3276>. Analysis of this protein sequence reveals the following:

```

Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1638 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

-1186-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CA312294 GB:899106 similar to transposon protein [Bacillus subtilis]
Identities = 84/291 (28%), Positives = 130/291 (46%), Gaps = 6/291 (2%)

5 Query: 6 MLDYIAVTIKGLAPDDVIEKILILPKDKFVLNEMGINKYQRHYSFSEKIKVYPNKOWQSKM 65
M+D+ V+ K D +E+L L KD + G Y Y 1KV+++ ++
Sbjct: 31 MVDYIRVSFKTHDVDRITIEVLHLSKDFMTEKQSGFYGVGYVELDYIKVFPYSAPODMR- 89

10 Query: 66 GVFIELRCGCGCRQYREYEMNVMNVTIMKRISSECHSNVTRLDIANDIFDSSLGVLPLYS 125
GV IE+ QCGCRQ+E +E H + + + TR D+A D S+P +
Sbjct: 90 GVLIEMSGCGCRQFESFLECRKKTWIDFPQDCMQGGSFTRFDLADOKKTYFPIELPK 149

15 Query: 126 YCKQLCISTAKTDFDIEKSLLENKEKVGEMFTIGVRCQQW-CVYNKILEQKLDQELPN 184
+K CIS + D++ L+G G + G + ++ C Y K EQ +P
Sbjct: 150 KAKGECISRFKESDFNGSFDLSDGITGTTIYFSGSKSEAYLCFYEKVQEAKNIPI 209

20 Query: 185 TPL-SWTRAEKRWQSKANLLAKQIKERPLKRIEYFVINGHYRFPVSPDKDGSNRWRKT 243
L W R ZLR E+A + + + L I ++IN + RFV D++ R KT
Sbjct: 210 EELADNRYELRLKNERAQAIDALLKTKDILIAMQIINNYVRFD-ADENITREHKT 268

Query: 244 VGMNDYLETQKQKTVLSVKRTKPLKRSKQVTEKQVSRILGLYVAKASH 294
+W+D++ + L VK K ++S W + T+ V +A+ H
Sbjct: 269 SLFNSDFIGDVGRLPLYVKQDFQYQKSRWLRNSCAPTM--KMLLEADH 317

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1066

30 A DNA sequence (GBSx1139) was identified in *S.agalactiae* <SEQ ID 3277> which encodes the amino acid sequence <SEQ ID 3278>. This protein is predicted to be integrase. Analysis of this protein sequence reveals the following:

Possible site: 58
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1914 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database.

>GP:CB970622 GB:AJ243106 integrase [Streptococcus thermophilus]
Identities = 135/474 (28%), Positives = 233/474 (48%), Gaps = 66/474 (14%)

45 Query: 20 KAGNVLVKEMRFTPHPTIKSKKKVLTGASKGWTYKATPSKLPSPGKERLLVSDIKNT 79
K G + VKF F + + T K ++ L S W+T +KK +GK +L S
Sbjct: 19 KTGIEVKFRTYFNNQLINK-RREILSD-----WYTIIV--NKKDTGKIKI--SPQIKA 67

Query: 80 QLITQVTPQELNLDVDDYAEIEMGIEKPKAKKLTLEIEAKFPDKDGNPYGAKFAWH-- 136
+ + + ++ KK+ ++ ++ K +TL+E+ + WH
50 Sbjct: 68 ITHKELQKANKVYZELRTIL-----LEKSDITLDEV-----WNEPHNER 108

Query: 137 -ERVKPANNLTAKTRVTIYKRYIEPHFDTIRMSITKEAFMTDEIQNLIN-----ASSMENAR 190
ER A TL Y + I + SI K + I+NI++ + A+
Sbjct: 109 VERQLVAPKTLAGEGGRYRNHTEQIP-KNSILK-NIPSSILKNLNDLNPYIGNHRLAQ 166

55 Query: 191 NLHIYLMKIFDWSVBNKQITLTQDPIASNKVRRVLTKESEQDK- KREDAEKYLEASV 249
+ L I+ + + + I+ Q+P+ + R+ L S+E D+ K+ DI ++YLE+ E+
Sbjct: 167 GVKSDLTSTYKFAILDHYISPDQPMFPIYSIGKGL--SOELRLKSDSDYQYLSKEL 224

-1187-

Query: 250 NIVLRLEISSWINRFDNQLIADVLRLMIFLTQMRPSEVLGLNEDMLDPEKKWIKVHMQRASK 309
 VL ++ + N+ A + + LTQMR EVLGI, E+ +DF K V RA+
 Sbjct: 225 KEVLSIVRKY-----NHCYARIFWEPQALITQMRIGEVLIHKEEAIDMNNTASVIRIRATH 279

5 Query: 310 NKSDMMALALDEKERYRADLKTESVRTIPMSPEVEKILRHVYIDRNKFPQAFSPITYQD 369
 + + + + Y ++K +S R +S +IL+ I+ N +F+D Y+D
 Sbjct: 280 GQASE-----DSYECNVNQLQSYRNQLSKRAIEILKEIRLHQHIRFNDPYKD 329

10 Query: 370 LGYLPRTTYIRAGMRQGSPLVHNKLSQPLRGSSQSASAKYKKGKPKYK---DIDSPLEFG 426
 G+PT I + G+PL+++L+ FL SS++ K N+ G P + DID+ L F
 Sbjct: 330 NGWIFTSKSIHKPDYKGTPLHYSLVNNPL--NSENKGLNRN-GNPRAGIDIDNLEPK 386

Query: 427 RPIHVFPHMFHSPFISIMASEGIDLPITREPVGHSEDSKEIERVYLHVIVKQKQD 480
 + H+ H+FRH+ IS +A +G+ L I++ VGHIS S+ + +YLH+ KK KD
 15 Sbjct: 387 K--HITTHIFRHTHLSFLAQGVPLBAIQDVRVGHSGSSR-VTEIYLHITKTKKD 437

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3279> which encodes the amino acid sequence <SEQ ID 3280>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5203 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 82/357 (22%), Positives = 155/357 (42%), Gaps = 52/357 (14%)

30 Query: 135 WHERVKRANVTLTKRTVITYKRIEVPFDFRMSITKFAFMTIEQLNDNA--SMHMRANL 192
 W K +T + R + D + I K T +Q+I+ S +
 Sbjct: 73 WEHQKSLKSTSVSLDFRIRKELNLDPEVMIKAIT--TKYLSIIDKIPGSYDRKRA 130

35 Query: 193 HIYLCMIFDMSVENSQITLTCQPIASNKVRRVLTKEEQQKQKEDIAKYLEASEVHV 252
 LK FD+++ + + + +P+ S + + + V T K ED+A+K+LE E+
 Sbjct: 131 RQLLKQTFDYAIALEYVSI--NPVISTQLAKPVKTI-----KDFEDVAKFLKDEKEL-- 181

40 Query: 253 LRLIESWINRFDNQLIADVLRLMIFLTQMRPSEVLGLNEDMLDPEKKWIKVHMQRASKKES 312
 RL++ R + +A + + +L GR E L + D + + I++H
 Sbjct: 182 -RLDEMYRRKGSIOAYLAEPMSLNGCRIGALAIQPD--NIKNDIIEH----- 229

45 Query: 313 DDMMALALDEKERYRADLKTESVRTIPMSPEVEKILRHVYIDRNKFPQAFSPITYQD 372
 ++ + + + + NT SR ++ ++I++ + N + +P Y+D+GY
 Sbjct: 230 -GTLDYTSNGYRNAIKITTPVMSWRETLITKREKIIQDILKINALEKRVNPNYKIMGY 288

Query: 373 LFTRTYIRAGNRQGSPLVHNKLSQPLRGSSQSASAKYKKGKPKYKIDSFLDFRPIHVI 432
 +F +R G P+ N L+ +R NK+ KP + +
 Sbjct: 289 IFI-----SRNGVPIDQNALMTSIRAA-----NKRLEKPIQK-----ELT 323

50 Query: 433 PHMFHSPFISIMASEGIDLPITREPVGHSEDSKEIERVYLHVIVKQKQDVRGAVEKL 489
 H+FRH+ +S +A + L TI + VGH+ DSK +++Y HV K K+ + +L
 Sbjct: 324 SHIFRHTLVSLKLAENKVPKLTIMDVRGHA-DSKTTQOITTEVTKSMKNKNEVDILNRL 379

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1067

A DNA sequence (GBSx1140) was identified in *S.agalactiae* <SEQ ID 3281> which encodes the amino acid sequence <SEQ ID 3282>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

-1188-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3023(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10277> which encodes amino acid sequence <SEQ ID 10278> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AB64982 GB:U43834 Ydr540cp [Saccharomyces cerevisiae]
 Identities = 88/170 (51%), Positives = 117/170 (68%), Gaps = 3/170 (1%)

Query: 36 MPTYSQWNLKEEVLKSYKKYIARFNDIPEKIKLRIDEVDRTPAENILAYQGVWTLILK 95
 MR Y+ K ELKEE+ K Y+KY AEF I E KD +++ VDRTP+ENL+YQ+GW L+L+
 Sbjct: 1 MREITSKELKEEIEIKYKDYAEFETISESQDEKVEVDRTPSNLSTQLQWNLILE 60

Query: 96 WESDEQSGLEVKTPPTETFKWNLGELYCHFTYASLTIKELTAQLAENVDAGRMIDEN 155
 WE+ E +G V+TP +QWN LG LYQ F + Y +IKE A+L + V+ + I ++
 Sbjct: 61 WRAKEIAGYNVETPAQYIOWNLGGLYQSFKYKGIYS:KEQRAKLEAVNEVYKWI STL 120

Query: 156 SDEVLFKPHMHWADGATIGAVWVEVTKFIHINTVAPPTGTFKIRKWKV 205
 SD+ LF+ R W AT A+W VTK+IHINTVAPP FR KIRKWK++
 Sbjct: 121 SDDSLFQAGNRKN---ATTYQWVPTVYKWIHINTVAPPTMFRGKIRKWKRL 167

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1068

A DNA sequence (GBSx1141) was identified in *S.agalactiae* <SEQ ID 3283> which encodes the amino acid sequence <SEQ ID 3284>. This protein is predicted to be 50S ribosomal protein subunit L33-related protein. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5420(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AB66692 GB:U89998 50S ribosomal protein subunit L33
 [Lactococcus lactis subsp. cremoris]
 Identities = 43/49 (87%), Positives = 46/49 (93%)

Query: 1 MRVNITLHKESGERLYLTCKNRMTFPDRILQKRYSPKLRKHVVTFEVK 49
 MRVNITLHKESGERLYLT ENKRNFTPD+L+LKRYK KLRKH+V EVK
 Sbjct: 1 MRVNITLHKESGERLYLTCKNRMTFPDKLELKRYSKLRKHVIFKEVK 49

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3285> which encodes the amino acid sequence <SEQ ID 3286>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5394(Affirmative) < succ>

-1189-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 48/49 (97%), Positives = 48/49 (97%)
 Query: 1 MRVNITLLEHKESGERLYLT/SKNKRNTPORLQLKKYSPKLRKHVVFTEVK 49
 MRVNITLLEHKESGERLYLT/SKNKRNTPORLQLKKYSPKLRKHV FTEVK
 Sbjct: 1 MRVNITLLEHKESGERLYLT/SKNKRNTPORLQLKKYSPKLRKHVVFTEVK 49

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1069

A DNA sequence (GBSx1142) was identified in *S.agalactiae* <SEQ ID 3287> which encodes the amino acid sequence <SEQ ID 3288>. This protein is predicted to be 50S ribosomal protein subunit L32-related protein. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3577(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB66691 GB:U89998 50S ribosomal protein subunit L32
 [Lactococcus lactis subsp. cremoris]
 Identities = 44/53 (83%), Positives = 48/53 (90%)
 Query: 1 MAKPARHTSKAKENKRRTHYKLTAPSVQFDETTG DYSGHRVSLKGYKGRKI 53
 MA PARHTS AKANRTHYKLTAP+V FDETTGDY SHRVSLKGYKGRK+
 Sbjct: 1 MAVPARHTSSAKKRRRTHYKLTAPTFTFDETTGDYRSHRVSLKGYKGRKV 53

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3289> which encodes the amino acid sequence <SEQ ID 3290>. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5148(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 38/39 (97%), Positives = 39/39 (99%)
 Query: 22 LTAPSVQFDETTG DYSGHRVSLKGYKGRKIAKANEAK 60
 +TAPSVQFDETTG DYSGHRVSLKGYKGRKIAKANEAK
 Sbjct: 1 NTAPSVQFDETTG DYSGHRVSLKGYKGRKIAKANEAK 39

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1070

A DNA sequence (GBSx1144) was identified in *S. agalactiae* <SEQ ID 3291> which encodes the amino acid sequence <SEQ ID 3292>. This protein is predicted to be histidyl-tRNA synthetase (hisS). Analysis of this protein sequence reveals the following:

```

5      Possible site: 32
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.4357 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10275> which encodes amino acid sequence <SEQ ID 10276> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA78919 GBS:Z17214 histidine--tRNA ligase [Streptococcus
      equisimilis]
      Identities = 327/404 (80%), Positives = 361/404 (88%)

20 Query: 32 WQYVENVIRNLFKQYHYDEIRTPMFERYEVISRSVGDTTDIVIKEMYDFHDKGORHITLR 91
      WQYVE V R FKQYHY RIRTPMFERYEVISRSVGDTTDIVIKEMYDF+DKGRHITLR
      Sbjct: 1 WQYVSGVARETFKQYHYGEIRTPMFERYEVISRSVGDTTDIVIKEMYDFDKGRHITLR 60

25 Query: 92 PEGTAPVVRYSYVENKLFAPFVQKPTMYIYGSMFRYERLPQAGRLREFHQGVGVECPGSGNP 151
      PEGTAPVVRYSYVENKLFAPFVQKPTMYIYGSMFRYERLPQAGRLREFHQGVGVECPGSG NP
      Sbjct: 61 PEGTAPVVRYSYVENKLFAPFVQKPTMYIYGSMFRYERLPQAGRLREFHQGVGVECPGSGANP 120

30 Query: 152 ATDVETIAMGHHLFEDIGIKNVKLIHNSLGNPESRQAYRQALIDYLTPIREQLSKDSQRR 211
      ATDVETIAM+HLFE LGIK V LIHNSLGN SR AYRQALIDYL+P+R+ LSKDSQRR
      Sbjct: 121 ATDVETIAMAYHLFERLGIGVTLHNSLGNASRAATRQALIDYLSMRDTLSKDSQRR 180

35 Query: 212 LNNENPLRVLDKSKEPEDKLAIVANAPSILDYDESSQAHFDVANCHMLDALNIPYIIDTNMVR 271
      L+ENPLRVLDKSKE EDK+AV NAPSILDY DE SQAHFDV ML+AL IPI+IDTNMVR
      Sbjct: 181 LDENPLRVLDKSKEKEDIKLAIVANAPSILDYDESSQAHFDVAVRMLALAIPIVIDTNMVR 240

40 Query: 272 GLDYNNHTIFEFIETIEDNELTICAGGRYDGLVSYFGSGPTAFGFGGLGLERILLILLDKQ 331
      GLDYNNHTIFEFIETI++ +ELTICAGGRYDGLV YFGSG TP FGFGGLGLERILLILLDKQ
      Sbjct: 241 GLDYNNHTIFEFIETIVDSELTICAGGRYDGLVETFGSGPTAFGFGGLGLERILLILLDKQ 300

45 Query: 332 GISLPIENTIDLYIAVLGSBANLAALDLAQSIRHQFKYKVERDYLGRKIKAQFKAADTF 391
      G+ LP+E +D+YIAVLG++AN+AL L Q+IR QSF VERDYLGRKIKAQFKAADTF A
      Sbjct: 301 GVSLPVEBGLDVYIAVLGADANVAALALTQAIRRQGFVERDYLGRKIKAQFKAADTFKA 360

Query: 392 KVMITLGSSEVDSKEVGLKRNQYTRQGVKVSFEMIKTDFSSVLQ 435
      KV++TLG SE+ + + LK+NQYTRQV+ VEF+ I+IDF+S+ +
      Sbjct: 361 KVVITLGSSEIKAGQAVLKRNQYTRQGVKVSFEMIKTDFSSVLQ 404

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3293> which encodes the amino acid sequence <SEQ ID 3294>. Analysis of this protein sequence reveals the following:

```

50      Possible site: 27
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
55      bacterial cytoplasm --- Certainty=0.3183 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 339/424 (79%), Positives = 367/424 (90%)

Query: 13 NKLQKPKGTQDILPGESAKWCYVENVIRNLFKQYHYDEIRTPMFHYEVISBSVGDTTDI 72
 NKLQKPKGTQDILPG+AKNQYVE+V R+ F QY+Y EIRTPMFHYEVISBSVGDTTDI
 Sbjct: 1 NKLQKPKGTQDILGDAKNQYVESVARDTFSQNYGERTMPFHYEVISBSVGDTTDI 60

Query: 73 VTKEHYDFHDKGRHITLRFEGTAPVVRSYVENKLFAPFVQKPTWYVIGSMFRYERPOA 132
 VTKEHYDF+DKGRHITLRFEGTAPVVRSYVENKLFAPFVQK K+YVIGSMFRYERPOA
 Sbjct: 61 VTKEHYDFYDKGRHITLRFEGTAPVVRSYVENKLFAPFVQKPVKLYYIGSMFRYERPOA 120

Query: 133 GLRLEFHQGVGECFG+NPATDVETIAM+HLFE LGIK+V LHLNSLG+PESR AYNQA 192
 GLRLEFHQ+GVECFG+ NPATDVETIAM +HLFE LGIK+V LHLNSLG+PESR AYNQA
 Sbjct: 121 GLRLEFHQIGVECFGAAAMPATDVETIAMAYHLFEKILGIDVTLHLNSLGSPESRAAYRQA 180

Query: 193 LIDYLTPIREQLSKDSQKRLNENPLRVLDSKEPEDKLAVENAPSILOYLDESSQAHPAV 252
 LIDYLTPI+R+QLSKDSQKRL+ENPLRVLDSKE EDKLAVE APSILOYLDE SQAHP+AV
 Sbjct: 161 LIDYLTPIHFDQLSKDSQKRLDENPLRVLDSKEPEDKLAVEKAPSILOYLDESSQAHPAV 240

Query: 253 CMLDALNIPYIIDTMRVGLDYSHHTIFEFITIEDNELTICAGRYDGLVSYFGGPEF 312
 ML+AL+IPY+IDTMRVGLDY+HTIFEFIT +E ++LITICAGRYD LV YFGGPEF
 Sbjct: 241 KDLMLDALNIPYIIDTMRVGLDYSHHTIFEFITSVESDLTICAGRYDSLGVYFGGPEF 300

Query: 313 PPGFGLGLERLL+I+KQGI+LPIE +D+Y+AVLG AN AL+L QHIR QGF ER 372
 PPGFGLGLERLL+I+KQGI+LPIE +D+Y+AVLG AN AL+L QHIR QGF ER
 Sbjct: 301 PPGFGLGLERLLMIIEKQGITLPIETEMDYTLAVLGDGANSKALELQAIRQGFTRER 360

Query: 373 DYLGKRIKAQFESADTFKAKVMTLGESEVDSEKVGKIQNTQGVKVSFENIKTDFP 432
 DYLGKRIKAQFESADTF AK++MTLG SEV++ + +KRN++RQGV+VSFE++ TP+P++
 Sbjct: 361 DYLGKRIKAQFESADTFKAKVMTLGESEVDSEKGVKIQNRSGVEVSFEMTNPANI 420

Query: 433 LKQL 436
 +QL
 Sbjct: 421 SEQL 424

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1071

A DNA sequence (GBSx1145) was identified in *S. agalactiae* <SEQ ID 3295> which encodes the amino acid sequence <SEQ ID 3296>. This protein is predicted to be aspartyl-tRNA synthetase (aspS). Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.5124 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10273> which encodes amino acid sequence <SEQ ID 10274> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GF:CB14714 GB:Z99118 aspartyl-tRNA synthetase [Bacillus subtilis]
 Identities = 339/585 (57%), Positives = 432/585 (72%), Gaps = 9/585 (1%)

Query: 20 RSMYAGRVSRSHIGTISITLKGWVGRKRDGLGLIFIDLRDRGIMQLNIPESVSASVMAT 79
 R+ Y G + + IG S+TLGWV +RRDLGLIFIDLRDR GI+Q+V NP+ VS +A
 Sbjct: 4 RTTYCGDITEKAIQESVTLKGWVQGRDLGLIFIDLRDRGTIGVQVVPNPD-VSKRALAI 62

Query: 80 AESLRSEPFVIVSGVVIAREQA--NDNLPTGEVELKVOELSIINTSKITFPRIKDGE-A 136
 AE +R+E+V+++ G V ARE+ N NL TG +E+ +++IN +KT PF I D E

-1192-

Sbjct: 63 AEGIRNEYVLIDQKKVAREBETVNPPLKTAIRIHADGVNVLNAAKTPPFAISDQAEV 122
 Query: 137 NDTFRMYRYLRLRRPEMLNFPLRAKVTHSIRNYLDNLEFIDVETPMLTKSTPEGARDY 196
 ++D R++RYLRLRRP M + +LR VT ++R++LD P+D+RTP+LT STPEGARDY
 Sbjct: 123 SEDVRLKIRYLDLRRPAMPQITMQLRHNVTKAVRSPLENGFLDITETPILGSTPEGARDY 182
 Query: 197 LVPSRVNQHGFYALPQSPQITKQLLMNAGFIRYQIVKCFRDEDLRGDRQPEPTQVLETS 256
 LVPSRV++G FYALPQSPQ+ KQLM +G +RYQI +CFRDEDLR DRQPEPTQ+D+E
 Sbjct: 183 LVPSRVHSEFYALPQSPQQLQLMVGSIERYQIARCFRDEDLRGDRQPEPTQIDTEM 242
 Query: 257 FLSQDEIQDIVEGMIAKVMKDTKGLVSLPFFRMAYDDAMNNGSDKPDTRFMDLLQLD 316
 FLS++I+ +E M+AKVM+TRG E+ LP PRM YD+AMN YGSDKPDTRFMDLL D+
 Sbjct: 243 SFMSQEDIMSLAKEMMAKVMRETNGERLQLPLPRPTIDRMMNKYGSCKPDTRFMDLLTDV 302
 Query: 317 TRIVKEVDKVFSEA---SVVRAIVVDKADKYSRNIDKLTALRQYGAQGLAWLKYA 372
 ++IVK+ +FKVFS A VVKAI VK A YSRK+ID L A YGAKGLAW+K
 Sbjct: 303 SDIVKDTBFKVPSSAVANGGVVKAIVKVGAGDYSRSDIDALGAFANNGAKGLAWKVE 362
 Query: 373 DNTISGPKAFKL-TAIGRLTALQLENDLILFVADSLVANETIGALATRIAKELDI 431
 ++ GP+AKF ++L EAL DL+LF AD KV +GLALR ++ KE LI
 Sbjct: 363 ADGVKGPITAKFPDESEKQKLI EALDAEGDULLPGADQFEVVAASGLALRLKLGKRGGLI 422
 Query: 432 DYSKFNFLVVDVMPFMSSEEGRYMSAHHPFTLPTAETAHELBDLAKVRAVAYDIVLN 491
 D FNLFL+DWP+ E EGRG+ +AHHPFT+P E ++E ++A AYD+VLN
 Sbjct: 423 DEKLFNFLVVDVMPLEHDPBEGRFYAAHHPFTMPFVRDLLELITAPEMKQAQYDVLN 482
 Query: 492 GYELGGGSLIRINQKDTQSRMFKALGFSAESAQCGPFLLEAMDYGFPFHGGLAIGLDRFV 551
 GYELGGGSLRI +KD QE+MF LGFS E A EQPGFLLEA +YG FPHGG+A+GLDR V
 Sbjct: 483 GYELGGGSLIRIEKDIQKQIFALLGFSFREAABQPGFLLEAFYBGFPFHGGLAIGLDRV 542
 Query: 552 MLIAGKNIREVIAFPKKNKASDPMTPQAPSLSVSEQQLESLTVE 596
 MLIAG+ N+R+ IAFPK AS MT+AP VS+ QL+EL L+++
 Sbjct: 543 MLIAGRTNLRDTLAFPKTASASCLMTEADGEVSDAQLEHLHLK 587

- 35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3297> which encodes the amino acid sequence <SEQ ID 3298>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have an uncleavable N-term signal seq

40 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 45 An alignment of the GAS and GBS proteins is shown below.

Identities = 495/582 (85%), Positives = 538/582 (92%)

Query: 18 MKRSMYAGRVREHIGTSITLKGWRRRRDLGGLIFIDLRDRBGIMQLVINPEEVSASVM 77
 MKRSMYAGRVREHIGT+ITLKGWV RRRDLGGLIFIDLRDRBG+MQLVINPEEVS+ VM
 Sbjct: 18 MKRSMYAGRVREHIGT+ITLKGWRRRRDLGGLIFIDLRDRBGVQMLVINPEEVSIVM 77
 Query: 76 ATASRLRSSEVIEVSGVVTAREQANDMLPTGEVELKVOELSLANTKTPTRFRIKDIQAN 137
 ATAE LRSK+VIEV G V AR+QAND L TG VELKV L+ILNT+KTPTRFRIK D+
 Sbjct: 76 ATASRLRSSEVIEVSGFVEARQANDKLATGMVELKVSALTILANTKTPTRFRIKDVESV 137
 Query: 138 DTRLRMYRYLDLRRPEMLNFPLRAKVTHSIRNYLDNLEFIDVETPMLTKSTPEGARDYL 197
 DTR+RYRYLDLRRPEMLNFPLRAKVTHSIRNYLD+LEFIDVETPMLTKSTPEGARDYL
 Sbjct: 138 DTRLRMYRYLDLRRPEMLNFPLRAKVTHSIRNYLDLEFIDVETPMLTKSTPEGARDYL 197
 Query: 198 VPSRVNQHGFYALPQSPQITKQLLMNAGFIRYQIVKCFRDEDLRGDRQPEPTQVLETS 257
 VPSRV+CHGFYALPQSPQITKQLLMNAGFIRYQIVKCFRDEDLRGDRQPEPTQVLETS
 Sbjct: 198 VPSRVNQHGFYALPQSPQITKQLLMNAGFIRYQIVKCFRDEDLRGDRQPEPTQVLETS 257
 Query: 258 FLSQDEIQDIVEGMIAKVMKDTKGLVSLPFFRMAYDDAMNNGSDKPDTRFMDLLQLD 317
 FLS+QDEIQDIVEGMIAKVMK+TK ++V+LFFRM+YD AMN+YGSCKPDTRF+MLQLD

-1193-

Sbjct: 256 FLSEKIQDIVEGMLAKVMKETEIDVTLPPFRMSYDVAMNSYGSKDPUTRFEMLLQDLT 317

Query: 316 RIVKEVDPKVFSEASVVKATVVKDKADKYSRKNIDKLTIAKQYAGRLAWLKAYDNTIS 377
VK DPKVFSEA VKATVVK AD+YSRK+IDKLT AKQ+GAKGLAW+K D ++

5 Sbjct: 316 VTVKGNDFKVFSEAPVKAIVVKNADRYSRKIDKLTFAKQPCAGRLAWVKVDQQLA 377

Query: 378 GPVAKFLTAIRGRITRAIQLENNDLILFVADSLRVANETLGAITRIAKKLEIDYSKPN 437
GPVAKFLTAIR L+ L+L NDL+LFVAD+LRVAN TLGALR RIAR+L+ID 8+FN

10 Sbjct: 378 GPVAKFLTAIRTELSSQLKLAENDLVLFAVITLEVANITLGAIRNRIAKLDMDIQSQFN 437

Query: 438 FLNVVWVWPFMSSESGRYMSAHPFTLPTAEHTAHEGLDLAKVRNAVYDIVANGVLEGG 497
FLNVVWVWPFMSSESGRYMSAHPFTLPT E+AHLEGLDLAKVRA+AYDIVANGVLEGG

15 Sbjct: 438 FLNVVWVWPFMSSESGRYMSAHPFTLPTPSAHLKGLDLAKVRAIYDIVANGVLEGG 497

Query: 498 GSLRINQKDTQERMFKALQPSAESAQDQPFLLRMDYGFPHGGLAIGLDRFVMLLAGK 557
GSLRINOK+ CERMFKALGF+A+ A +CPGFLLEMDYGFPHGGLAIGLDRFVMLLAGK

20 Sbjct: 498 GSLRINQKQERMFKALGFADENDQPFLLRMDYGFPHGGLAIGLDRFVMLLAGK 557

Query: 558 DNIREVIAPFNKNKASDPMWQAPSLVSEQLLESLGIVSEY 599
DNIREVIAPFNKNKASDPMWQAPSLVSE QLESLG +ES++

20 Sbjct: 558 DNIREVIAPFNKNKASDPMWQAPSLVSEQLLESLGQIESHD 599

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1072

A DNA sequence (GBSx1146) was identified in *S. agalactiae* <SEQ ID 3299> which encodes the amino acid sequence <SEQ ID 3300>. Analysis of this protein sequence reveals the following:

Possible site: 54

>>> Seems to have no N-terminal signal sequence

30 INTEGRAL Likelihood = -8.44 Transmembrane 186 - 202 (182 - 205)
INTEGRAL Likelihood = -5.68 Transmembrane 88 - 104 (86 - 106)
INTEGRAL Likelihood = -3.40 Transmembrane 115 - 131 (112 - 132)
INTEGRAL Likelihood = -2.13 Transmembrane 141 - 157 (141 - 157)
INTEGRAL Likelihood = -0.96 Transmembrane 43 - 59 (43 - 59)

35 ----- Final Results -----

bacterial membrane --- Certainty=0.4376 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

40

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12952 GB:Z99109 alternate gene name: yuxa-similar to
hypothetical proteins [Bacillus subtilis]

45 Identities = 104/275 (37%), Positives = 161/275 (58%), Gaps = 1/275 (0%)

Query: 39 EKISASLIVGILSVAVNFFPQPHVYSSCATGLAQVISAVSKHWFSFEIVALAFYAIN 98
+K+ ++ +L+ +N F P VY+SG TC+AQ+SE+V + F I +N
Sbjct: 7 KKLILIVIGALINAGLNLFLIPADVYASGPTGVAQLSSVDDQZAFFYISTGTLFLIAN 66

50 Query: 99 IPILLISWRKIKHKPTIPTFTITVYSSPIQAMPQITLTDPLINAIPOGLINGMGVQFS 158
IP+ IL W K+G FT+++ +V +++++F+ ++P+ +L D L+N+PGQ+I G+G +
Sbjct: 67 IPVIGLWLVKVGKSFVYISLVALITLPMGILPETSLSHDILLNAVFGGVISAVGIGLT 126

55 Query: 159 PKSRISGQDTDISLTIRKTRCDRVGSIISFINGIILLFACLLPGWIKALYSMTTIPVSS 218
K S+GG DI+++ + K + VQ+ FI+NGII+L AGLL G+ ALX++VT+++V++
Sbjct: 127 LKYGASTGGLDIVAMVLAKWKDKPVGTYFTPIINGIIITLGLLQWEXALYTVTVTVTT 186

60 Query: 219 RVIDAIFTQKQKQMIIVTSKFPYCVIKRHRDLHRGVTCINDABTYTHNEKGAVLITIL 278
RV DAI T+ K+ AMIVT K + + I+ + RG+T + A+G + +E+K +I +T
Sbjct: 187 RVIDAHTRHMKLTAMIVTKKAEILKEALYKGVKVSITTV-PKGAFTNSEQEKMIIIVT 245

Query: 279 RERFSDPKYLLMLKADPKAPVSAENVHIIIGRFVDD 313

-1194-

RE D + + + DPKAF + + + I G F D
 Sbjct: 246 RYELYDLRKIVKEVDPKAFINIVQTIGIFGFRKD 280

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3301> which encodes the amino acid
 5 sequence <SEQ ID 3302>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence
 10 INTEGRAL Likelihood = -5.47 Transmembrane 87 - 103 (86 - 106)
 INTEGRAL Likelihood = -4.94 Transmembrane 185 - 201 (182 - 203)
 INTEGRAL Likelihood = -1.59 Transmembrane 114 - 130 (113 - 130)
 INTEGRAL Likelihood = -1.12 Transmembrane 42 - 58 (42 - 58)
 INTEGRAL Likelihood = -0.32 Transmembrane 140 - 156 (140 - 156)
 ----- Final Results -----
 15 bacterial membrane --- Certainty=0.3187 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

20 >GP:CAA66894 GB:X98238orf2 [Lactobacillus sakei]
 Identities = 105/280 (37%), Positives = 180/280 (63%), Gaps = 7/280 (2%)
 Query: 37 AEKISALLYGILSSIAVNFPPQPGHVSAGATGLAQVFSAL-SHRLLYDPIAPAFYL 95
 +++I +++VG L++++VN F P YSSG TG+AQ+ +AL SH LG +A ++
 25 Sbjct: 8 SKRIVIAVWGFLAAVSNVLFIPAKTYSSGVTGVAQLLTALVSH--LGSSLSVAALVFI 65
 Query: 96 INIPLLILAWYKIGHQPTITFTITVSMSSFIQIMPOVT--LTDTPLINAFGLVMGMS 153
 +N+PL+LAW+KI HQ+ IF+ + V S F++I+P + T+ A+FGS+G+G
 Sbjct: 66 LNVPLLVLANFKINHQAIFSI VAVFTSVIFLKLIPVPQPIILTERFAGALFGALIGLG 125
 30 Query: 154 IGTGLKSRISGGTIDIVSLTLRKTKGVDVGSILSLMVGAILAFAGILFGWQYALYSMVSI 213
 +G ++ S+GGT++ ++ TGR VGG++ +NG I+ AGI FGW ALYS+V I
 Sbjct: 126 VGLCFRAGPSFTGGTDVITVLGRITGRVAVENVMGILAAIGFFGGAALYSIVEX 185
 35 Query: 214 FVSSRVTDALFTKQKQMAITVISHPERVIMHKLHKGVTISINDAGTYTHGQKAVLI 273
 FVSS + D I+T+Q+K+ TI T PE + + +H G T + D G Y +++ +V++
 Sbjct: 186 FVSSLLMDVITYTQOKQVTVITFTKQPEALKKRMREFIH-GATEL-DGTGLYNGQTSVIM 243
 Query: 274 TILTCREYEPFKMLMLKTDPAFVSVAVNRIIGRFVEDD 313
 40 T+++ + K ++ DP AFV++ ++ GRF ++
 Sbjct: 244 TVVSKYDLTALKLVQDADPAFVNITQSTNMAGRFESNE 283

An alignment of the GAS and GBS proteins is shown below.

Identities = 239/311 (76%), Positives = 274/311 (87%)
 45 Query: 4 RRTPLEKKVYIISVWAKKPLGLHLTKLSISREKYAEKISASLLYGILSSVAVNFPPQPGH 63
 ++T +KRVKY+IS ARK GLH L+SISREKYAEKISASLLYGILSSVAVNFPPQPGH
 Sbjct: 3 KATTYKKVYVIRGAUKVGLHLALRSISREKYAEKISASLLYGILSSVAVNFPPQPGH 62
 50 Query: 64 VYSSGATGLAQVISAIVKWSFBEIPVAPAFYAINIPLLILAWYKIGHQPTITFTITVTV 123
 VYSSGATGLAQV SA-S ++ P-A AFY INIPLLILAW KIGH+PTITFTITV++
 Sbjct: 63 VYSSGATGLAQVFSALSRLGYDFPIAPAPYINIPLLILAWYKIGHQPTITFTITVSM 122
 55 Query: 124 SSFIQIMPQITLTDTPLINAFGLLNGAGVGSFSSKRSISSGGTIDISLTIRKTKGRDV 183
 SS FIQ+MPQ+TLTDTPLINAFGL+NG+G+G KRSISSGGTID+SLT+RK+TG+DV
 Sbjct: 123 SSFIQIMPQVTLTDTPLINAFGLVNGMGIGTGLKRSISSGGTIDVSLTRKTKGRDV 182
 Query: 184 GSISPIINGIILLFAGLFGWQYALYSMVITFVSSRVTDALFTKQKQMAITVISHPERV 243
 GS+S +NG IL FAG+LFGW+VALYSMV+IFVSSRVTDALFTKQKQMAITVISHPERV 242
 60 Sbjct: 183 GSLSLMVNGAILAFAGILFGWQYALYSMVSI FVSSRVTDALFTKQKQMAITVISHPERV 242
 Query: 244 IKRIHRLHKGVTISINDAGTYHKEKAVLITILTRKFSDFKMLKADPKAFVSVAVN 303
 I IH+ LHRGV INDAEGTY HE+KAVLITILT EE+ +PK+LMK DP+AFVSVAVN
 Sbjct: 243 IRIHKLHKGVTISINDAGTYHKEKAVLITILTCREYEPFKMLMLKTDPAFVSVAVN 302

-1195-

Query: 304 VHTIGRFVDD 314
 V IIGRFV+DD
 Sbjct: 303 VHTIGRFVDD 313

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1073

A DNA sequence (GBSx1147) was identified in *S. galactiae* <SEQ ID 3303> which encodes the amino acid sequence <SEQ ID 3304>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -3.72	Transmembrane	156 - 172 (156 - 174)
INTEGRAL	Likelihood = -3.03	Transmembrane	112 - 128 (110 - 129)
INTEGRAL	Likelihood = -2.34	Transmembrane	80 - 96 (79 - 96)
INTEGRAL	Likelihood = -1.49	Transmembrane	60 - 76 (58 - 76)

 ----- Final Results -----

bacterial membrane	---	Certainty=0.2487(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05397 GB:AP001512 unknown conserved protein (Bacillus halodurans)
 Identities = 113/278 (40%), Positives = 192/278 (68%), Gaps = 1/278 (0%)
 Query: 7 KTKIKETILLIAGFVALYTFGFKFMNANHLAEQGISGVILLIHALPGVNPALSSLLINIP 66
 + K K + I G A+++FG V FM N+LAEIG +G+TLI++ +F +NPA+++L+NLIP
 Sbjct: 4 RLKWNINPILLGSAIPSGFLVYFMNENLAEQGISGVTITLLYFMQINPAVNTNLVNLIP 63
 Query: 67 LFILGRIILGKKSLLTIYGTVLMSFFPMFQQIP-VITVFLKNDMLVAAGILAGTGS 125
 + ++G +ILG+ +L+ TI GTV +S F+ +Q+ + +PL +DM L A+ AG+ GTG
 Sbjct: 64 ILLIGWKILGRVTLIYTIIGTVSVSVFLPMFQKWKPNIDPLHDDMTLAAFGVFGVGL 123
 Query: 126 GLVFRYGATGGADIIGRIVERKSGIKLQTLFLFIDAIVLTSSLVYINQQMLTYTVASF 185
 G+VFR+G TTGG DII ++ G +G+T+ DA+V+ SSL+Y+N ++ +YTL+A F
 Sbjct: 124 GIVFRGGTGGVDIIAKLGFYLGWSMGKIMFMDAVVIASSLIYNTREAMTILLAVF 183
 Query: 186 VFSQVLTVNENGGYTVRGMIIITKESESAANTILHEINRGVTELRGGGAYSRRHDVLVY 245
 + ++V+ ++ Y+ + IIT+ +E+ A TIL E+ RG T L+G+G++G E ++LY
 Sbjct: 184 IAAKVIDPITQTAYSAAAFIISHTAIAIDTILKMERGATLKGKSGPTGTEKILYLC 243
 Query: 246 ALNPSVDRVKNESIMADLDFAFISVINVDEVISSDPKI 283
 + +E+ +K ++ +DP AF++V +VI F +
 Sbjct: 244 VVGRENILRLKSLVERIDFAFVTVNDVQVIGSGFTL 281

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3305> which encodes the amino acid sequence <SEQ ID 3306>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -5.15	Transmembrane	112 - 128 (109 - 130)
INTEGRAL	Likelihood = -2.34	Transmembrane	156 - 172 (156 - 174)
INTEGRAL	Likelihood = -1.81	Transmembrane	178 - 194 (177 - 194)
INTEGRAL	Likelihood = -1.65	Transmembrane	80 - 96 (79 - 96)
INTEGRAL	Likelihood = -0.37	Transmembrane	60 - 76 (59 - 76)

 ----- Final Results -----

bacterial membrane	---	Certainty=0.3060(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

-1196-

The protein has homology with the following sequences in the databases:

```
>GP:BA05397 GB:AP001512 unknown conserved protein [Bacillus halodurans]
Identities = 116/276 (42%), Positives = 182/276 (65%), Gaps = 1/276 (0%)

5   Query: 9   KLLKLFLIALGVAIYTFQFVNFNMANALAEQGVAGITLILHAHGFNPAYSSLLFNLELF 68
      K   +   I   LG   AI++FG   V   PNM   N   LAEGG   GITLIL+   F   INPA   ++L+   N+P+
      Sbjct: 6   KWNKIFVLIGSAIFSPGLVYFNMENNLAEQGFITGITLILYPMFQINPAVNLVNLNPL 65

10  Query: 69   IIGAKIFGKRSALAIITYGTIVLMSAFINMMQKVP-IRGLKNDMMMLVAVVAGLFSGIGSGI 127
      ++G   KI   G+   L+   TI   GTV   +S   F+   M+Q+   +++   L+   DM   L   A+   AG+F   G   G   GI
      Sbjct: 66   LIGWILGRVTLIYTTIIGTVSVFLEMFQWKFMDIPLHDMTLALFAGVFVFGIGGI 125

15  Query: 128   VFRYGATTGGTDIIGRIAEKFGAKLQCTLLLDVLDVLTASLTIVVDLKHMLYTLVASFVF 187
      VFR+G   TTGG   DII   ++   G   +G+T+   +   DA+V+   +SL   Y++   +   +YTL+A   F+
      Sbjct: 126   VFRPGGTTGGVDIAKLGFRYLQWSMGKTMFMDAVVIASSLIYLYNREMYTLLAVFIA 185

      Query: 168   SQMISVVQNGGYTIRGMIIITGHSEAAQAAILTEINRGVYTLKQGAYSGNDYNIMYVL 247
      +++I   +Q   Y+   +   II++H+EA   A   IL   E+   RG   T   LKG+G+++G   +   I+Y   +
      20  Sbjct: 166   AKVIDFIQCTAYSAKAAFIISHTETAIAADTLKEMERGATTLKNGKGSFTGTEKILYCVV 245

      Query: 246   NPTEVREVRKILAGLDPDAFISIIDVDEVISSDFKI 283
      E+   +K   ++   +DP   AF+++   DV   +VI   F   +
      25  Sbjct: 246   GRNELIRLKSVERIDPHAFVTNVDVQDVIGEGFTL 281
```

An alignment of the GAS and GBS proteins is shown below.

Identities = 206/286 (72%), Positives = 250/286 (87%)

```
30  Query: 5   DLTKIKETILIAFGVALYTFQFVKFMNHLAEQGISGVTLILHALGVNIPALSSLLIN 64
      D   TK+   +   LIA   GVA+YTFGPV   FMNAN   LABGG+G+TLI+HA   FG+NPA   SSSL   N
      Sbjct: 5   DKLTLLKLFLIALGVAIYTFQFVNFNMANALAEQGVAGITLILHAHGFNPAYSSLLFN 64

      Query: 65   IPLFILGARILGKKSLLLITYGTIVLMSFFPMFQQIIPVTVPLKNDMMMLVAVAGILACTG 124
      +PLFILGA+I   GK+SL   LTIYGTIVLAS   F+N   WQ++P+   +   L+NDMMMLVAV   AG+   +G   G
      35  Sbjct: 65   LPLFILGAKIFGKRSALAIITYGTIVLMSAFINMMQKVPFIELGENDMMMLVAVVAGLFSGIG 124

      Query: 125   SOLVFRYGATTGGADIIGRIAEKSGIKLQCTLLFIDAIVLTSSLVYNLQOMLYTLVAS 184
      SG+VFRYGATTGG   DIIGRI   BEK   G   KLGQTL   +DA+VLT+SL   Y++L+   MLYTLVAS
      Sbjct: 125   SGIVFRYGATTGGTDIIGRIAEKFGAKLQCTLLLDVLDVLTASLTIVVDLKHMLYTLVAS 184

40  Query: 185   FVFSQVLINVENGGYTVRGMIIITKSESAATILHEINRGVTLFGQGAYSGREHEDVLY 244
      FVFSQ+++   V+NGGYT+RGMIIITK   SE+AA   IL   EINRGVTL+L+GGGAYSG   ++++Y
      Sbjct: 185   FVFSQMSVVQNGGYTIRGMIIITGHSEAAQAAILTEINRGVYTLKQGAYSGNDYNIMY 244

45  Query: 245   VALNPSEVRDVKEIMADLDPDAFISVINVDEVISSDFKIRRRNYDK 290
      V   LNP+EV+VR   K   I+A   LDDPAFIS+I+VDEVISSDFKIRRRNYDK
      Sbjct: 245   VTLNPTEVREVRKILAGLDPDAFISIIDVDEVISSDFKIRRRNYDK 290
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1074

A DNA sequence (GBSx1148) was identified in *S.agalactiae* <SEQ ID 3307> which encodes the amino acid sequence <SEQ ID 3308>. This protein is predicted to be BacB protein. Analysis of this protein sequence reveals the following:

```
55  Possible site: 60
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4355 (Affirmative) < succ>
60  bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
```

-1197-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 5 >GP:BAAL1330 GB:D78257 BacB [Enterococcus faecalis]
Identities = 27/88 (30%), Positives = 48/88 (53%), Gaps = 1/88 (1%)
- Query: 1 MPSEKEILDALSKVYSEEVQADDDYFQAIPELASOLEKEGMN-SLLATKIDSLINQVLT 59
M ++E+LD LSK Y++ I + + +FE A +L N + K+ ++ ++Y+
10 Sbjct: 1 MDKQQLLELLSKAYNDPKINYSGLKDKLPFCARILTNINIGEVCKLSTINSEYLA 60
- Query: 60 THQDAPKSIIDLSRLVKTAKASHYKGT 87
H F+ PKSI +L + V + Y+G A
Sbjct: 61 RHIFEMPKSIIEQLKPFVTKEQKYRGA 88
- 15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3309> which encodes the amino acid sequence <SEQ ID 3310>. Analysis of this protein sequence reveals the following:
- Possible site: 27
>>> Seems to have no N-terminal signal sequence
- 20 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2712 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 25 An alignment of the GAS and GBS proteins is shown below.

Identities = 99/102 (97%), Positives = 102/102 (99%)

Query: 1 MPSEKEILDALSKVYSEEVQADDDYFQAIPELASOLEKEGMN-SLLATKIDSLINQVLT 60
MPSEKEILDALSKVYSE+VIQADDDYFQAIPELASOLEKEGMN-SLLATKIDSLINQVLT
30 Sbjct: 7 MPSEKEILDALSKVYSEVIQADDDYFQAIPELASOLEKEGMN-SLLATKIDSLINQVLT 66

Query: 61 HQDAPKSIIDLSRLVKTAKASHYKGTAKAISAIMLGSPLSGGPK 102
HQDAPKSIIDLSRLVKTAKASHYKGTAKAISAIMLGSPLSGGPK
Sbjct: 67 HQDAPKSIIDLSRLVKTAKASHYKGTAKAISAIMLGSPLSGGPK 108

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1075

- 40 A DNA sequence (GBSx1149) was identified in *S.agalactiae* <SEQ ID 3311> which encodes the amino acid sequence <SEQ ID 3312>. This protein is predicted to be ArgS (argS). Analysis of this protein sequence reveals the following:

Possible site: 57
>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2522 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 50 A related GBS nucleic acid sequence <SEQ ID 10271> which encodes amino acid sequence <SEQ ID 10272> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF86984 GB:AF282249 ArgS [Lactococcus lactis subsp. lactis]
Identities = 377/566 (66%), Positives = 464/566 (81%), Gaps = 5/566 (0%)

55

-1198-

Query: 12 MDTGILIASRIQKVVD-MEQSTILSILETFKNSMGDLAFAPFSLAKTLRKAPQIIASD 70
MD K L++ + + + I +++E PK+S +GDLATPAF LAKTLRK+POLIA +
Sbjct: 1 MDEKQLVSQLASAAIDGVLGVETQIAIIIEKPKSEDLGDLAPFQLAKTLRKSPQIIAGE 60

5 Query: 71 IAEQIKSDQFEKVEAVGFPYVNFFLDKAAISQVLKQVLSGSAVATQNIQEGGRNVAIDMS 130
IAS+I + FEKV AVGFPYVNFFLDK A +S+V++VL+G Y NIGBG HV IDMS
Sbjct: 61 IAEKIDTKGFVKVAVGFPYVNFFLDKNATASEVIREVLASGEHYGDANIGEGGNVPIDMS 120

10 Query: 133 SPNIKPPSIGHLRSTVIGDSLNI+FDKIGYHPVKINHLGDWKGQPGMLIVAYKKWGNEA 190
+PNIKPPSIGHLRSTVIGDSL+I +K+GY P+KINHLGDWKGQPG+LI AYKK+G+E
Sbjct: 121 APNIKPPSIGHLRSTVIGDSLAKIYKELGYQPIKINHLGDWKGQGLLIITAYKKGDEA 180

Query: 191 AVRAHPIDELLKLYVRINAEATDPSVDEEAREWPKLEANDPEATELWQVPROBSILEP 250
+ A+PIDELLKLYV+INAE+ D VDEE R+WF K+E D EA +W+P D SL+EP
Sbjct: 181 TITANPIDELLKLYVKINAEAKEDSEVDEERQWFLKMQQDEALRIWQVPSVGLIEP 240

15 Query: 251 NRIYDQNVVTFPSYNGEAFVNDVMDVLELLESHNLLVESKGQVNVLEKYGIEHPALIK 310
NR+Y ++ VTFD + GE+FY+DMD ++E LE+HMLL ESKGA +V+LEKY + +FALIK
Sbjct: 241 NRIYKGLGVTFDHPMGESFYSKMDAIVEDLENHMLLHESKGALIVLEKYNL-NFALIK 299

20 Query: 311 KSDGATL+YITRDAAALYRKRTYDPAKSYVVGNEQSAHPKQLKAVLKEBMDYDSDNDTH 370
K+DGTAT+YITRDLA A YRK+T+P KS+YVVG EQ+ HPKQLKAVLKE YDSDNDH H
Sbjct: 300 KTDGATL+YITRDLATAYRKRTYDPAKSYVVGGEQTNHPKQLKAVLKEAGYDSDNDVH 359

25 Query: 371 VPPGLVTKGGAALSTRKGNVILEPVTARAINRAASQIEAENFNLDKVKYQAQVGAJ 430
VPPG+VT+GG K STRKG+V+ LE + EA++RA QIEAENFN +K+VA+ QVCGA+
Sbjct: 360 VPPGNVTQGGKKPSTRKGHVLEMDALDEAVRAEKQIEAENFNLENKEEVAQVQVGAJ 419

30 Query: 431 KFYDILKTDRTNKGDFDLDEAMVSFEGSTGFPYQVYAHARIQSILKANKPSNSDNYSL-N 488
KFYDILKTRD NGYDFDL+ MVSFEGSTGFPYQVYAHARIQSILRKN N DN SL +
Sbjct: 420 KFYDILKTDNNKGDFDLDEAMVSFEGSTGFPYQVYAHARIQSILKANK-RKNVDNLSLWS 478

Query: 489 DVSEWEIILKLIQFPFIRVRAADNFPESIIAKFAINLAQCFNKYIYARTRILDEDAISRS 548
D EWEI+K +++FP I+ RAADN+BPESIIAK+AI+LAQ FNKYIYH RLL+DA+ R
Sbjct: 479 DRAEWEIIVKALKEFPFNIKRAADNYPESIIAKFAISLAQAFNKYIYARVRLIEDDAQLQGR 538

35 Query: 549 LALCYATATVLKESLRLLLGVDAFNM 574
LAL AT+ VLKE+LRLLV AP M
Sbjct: 539 LALISATSVLKEALRLLLGVAAFNM 564

40

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3313> which encodes the amino acid sequence <SEQ ID 3314>. Analysis of this protein sequence reveals the following:

Possible site: 46
>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1734 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50

An alignment of the GAS and GBS proteins is shown below.

Identities = 492/563 (87%), Positives = 526/563 (93%)

55 Query: 12 MDTGILIASRIQKVVDMEQSTILSILETFKNSMGDLAFAPFSLAKTLRKAPQIIASDI 71
MDTK LIASEI KVPV+EQ I +L+STFKNS MGDLAFAPFSLAK LRKAPQ+IAS++
Sbjct: 1 MDTGILIASRIKVPVLEQDAIFNLILETFKNSMGDLAFAPFSLAKVLRKAPQMIASEL 60

Query: 72 AEQIKSDQFEKVEAVGFPYVNFFLDKAAISQVLKQVLSGSAVATQNIQEGGRNVAIDMS 131
AEQI QFEKV AVGFPYVNFFLDKA ISSQVL+QV++ GS YA Q+ GGRNVAIDMS
Sbjct: 61 AEQIDESQFEKVVAVGFPYVNFFLDKAKISSQVLEQVITAGSDYAGQDQEGGRNVAIDMS 120

60 Query: 132 PNIKPPSIGHLRSTVIGDSLNI+FDKIGYHPVKINHLGDWKGQPGMLIVAYKKWGNEA 191
PNIKPPSIGHLRSTVIGDSL+I F K+GY PVKINHLGDWKGQPGMLIVAYKKWG+E A
Sbjct: 121 PNIKPPSIGHLRSTVIGDSLAIHFAKMGYKPVKINHLGDWKGQPGMLIVAYKKWGDEA 180

65